Deliverable D4.1

Validation Report

Point of Contact	Samuel Fricker
Institution	FHNW
E-mail	samuel.fricker@fhnw.ch
Phone	+41 79 196 9629

Project Acronym	GEIGER
Project Title	GEIGER Cybersecurity Counter
Grant Agreement No.	883588
Topic	H2020-SU-DS03
Project start date	1 June 2020
Dissemination level	Public
Due date	30 November 2021
Date of delivery	30 November 2021
Lead partner	ULEI
Contributing partners	FHNW, CLUJ IT, BBB, SRA, PHF, MI
Authors	Max van Haastrecht (ULEI), Marco Spruit (ULEI), Bettina Schneider (FHNW), Louis Baumgartner (FHNW), Stelian Brad (CLUJ-IT), Jürg Haller (BBB), Renata Säuberli (BBB), Tony van Oorschot (SRA), Bernd Remmele (PHF), Wissam Mallouli (MI)
Reviewers	Amedeo D'Arcangelo (KSP), Cristian Priboi (CERT-RO)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 883588 (GEIGER). The opinions expressed and arguments employed herein do not necessarily reflect the official views of the funding body.

Revision History

Version	Date	Author	Comment
0.1	1st October 2021	Max van Haastrecht (ULEI)	Table of contents
0.2	27th October 2021	Max van Haastrecht (ULEI), Stelian Brad (CLUJ-IT), Tony van Ooorschot (SRA), Bernd Remmele (PHF), Wissam Mallouli (MI)	First draft, including ULEI, CLUJ-IT, SRA, PHF, and MI
1.0	5th November 2021	Amedeo D'Arcangelo (KSP), Cristian Priboi (CERT-RO), Max van Haastrecht (ULEI)	Changes based on review of first draft
1.1	17th November 2021	Max van Haastrecht (ULEI), Bettina Schneider (FHNW)	Second draft, including alterations and additions by ULEI and FHNW
2.0	24th November 2021	Amedeo D'Arcangelo (KSP), Cristian Priboi (CERT-RO), Max van Haastrecht (ULEI)	Changes based on review of second draft
2.1	26th November 2021	Max van Haastrecht (ULEI), Marco Spruit (ULEI), Louis Baumgartner (FHNW), Stelian Brad (CLUJ-IT), Jürg Haller (BBB), Renata Säuberli (BBB), Tony van Oorschot (SRA), Wissam Mallouli (MI)	Final version, including alterations and additions by ULEI, FHNW, CLUJ-IT, BBB, SRA, and MI
2.2	39th November	Bettina Schneider (FHNW) Samuel Fricker (FHNW)	Quality check

Contents

Abbreviations, Participant short names and Glossary

Abbreviations

BPMN	Business Process Model and Notation
CERT	Computer Emergency Response Team
CJML	Customer Journey Modelling Language
CSD	Certified Security Defenders
CSIRT	Computer Security Incident Response Team
D	Deliverable
DPIA	Data Protection Impact Assessment
FA	Formative Assessment
GDPR	General data protection regulations[1].
IS	Information Systems
IUA	Interpretation and Use Argument
KPI	Key Performance Indicator
ME	Micro Enterprise
MOS	Mean Opinion Score
MoU	Memorandum of Understanding
MSE	Micro or Small-sized Enterprise
MVP	Minimum Viable Product
NCSC	The Swiss CERT National Cyber Security Centre
SME	Small or Medium-sized Enterprise
STS	Socio-Technical Systems
STS-FA	Socio-Technical System implementing Formative Assessment
STS-ml	Socio-Technical Systems Modelling Language[2]
UI	User Interface
UML	Unified Modelling Language
WP	Work Package

Participant short names

FHNW	Fachhochschule Nordwestschweiz
ULEI	Universiteit Leiden
TECH.EU	Fores Media Limited
KSP	Kaspersky Lab Italia Srl
PFH	Pädagogische Hochschule Freiburg

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679

[2] Dalpiaz, Fabiano, Elda Paja, and Paolo Giorgini. Security requirements engineering: designing secure socio-technical systems. MIT Press, 2016.

MI	Montimage EURL
KPMG	KPMG Somekh Chaikin Partnership
BBB	Berufsfachschule BBB Baden
ATOS	Atos IT Solutions and Services Iberia SL
SKV	Schweizerischer KMU Verband
HAAKO	haako GMBH
CERT-RO	Romanian National Cyber Security Directorate
CLUJ IT	Asociatia Cluj IT
E-ABO	e-abo Gmbh
SCB	Braintronix Srl
PT	Public Tender Srl
SRA	Samenwerkende Registeraccountants en Accountants-Administratieconsulenten
CL	Coiffure Loredana

Glossary

Argument-Based Validation The extent to which the interpretation and use argument (IUA) for a test is "coherent and complete and all of its inferences and assumptions are highly plausible."[3]

Construct Validity The extent to which a test measures what it claims to be measuring.[4]

Content Validity The extent to which "test items" are an appropriate "sample of a universe in which the investigator is interested."[4]

Criterion Validity The extent to which test scores serving as an operationalisation of a construct correlate with, or predict, a theoretical representation of the construct (i.e., the criterion).[4]

External Validity Answers the question: "To what extent can the findings be generalized to other populations and settings?"[5]

GEIGER Framework The GEIGER Toolbox deployed on an end-user's device (Section 4.4) and Cloud being the single back-end (Section 4.3). Together, the GEIGER Toolbox and the Cloud are the platform used to enable the GEIGER ecosystem (Section 3). The GEIGER Framework includes the GEIGER Indicator (Section 4.5) and can be tried using the GEIGER Testbed and Demo environment.

GEIGER Ecosystem A community of human, organisational, and software actors supported by the GEIGER Framework working together for helping MSEs to become secure and compliant with data protection regulations. The definition is based on the idea of software ecosystems proposed by Jansen et al. (2009).[6]

Internal Validity Answers the question: "Are there alternative causal explanations for the observed data?"[5]

[3] Kane, M. (2013). The argument-based approach to validation. *School Psychology Review*, 42(4), 448-457.

[4] Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, 52(4), 281.

[5] Mingers, J., & Standing, C. (2020). A framework for validating information systems research based on a pluralist account of truth and correctness. *Journal of the Association for Information Systems*, 21(1), 6.

[6] Jansen, S., Finkelstein, A., & Brinkkemper, S. (2009, May). A sense of community: A research agenda for software ecosystems. In *2009 31st International Conference on Software Engineering-Companion Volume* (pp. 187-190). IEEE.

Reliability Answers the question: “Do measures show stability across the units of observation?”[7]

Security Defenders A person educated to help an MSE to get protected (Deliverable D3.1).

Statistical (Conclusion) Validity Answers the question: Is our sampling approach sufficiently robust to rule out the possibility that results occurred by chance?[7] [8]

[7] Straub, D. W. (1989). Validating instruments in MIS research. *MIS quarterly*, 147-169.

[8] Mingers, J., & Standing, C. (2020). A framework for validating information systems research based on a pluralist account of t ruth and correctness. *Journal of the Association for Information Systems*, 21(1), 6.

List of Tables

List of Figures

Summary

The deliverable D4.1 Validation Report lays out the groundwork for the validation activities conducted in the GEIGER project. We connect the goals for the overall project to objectives during validation, where we aim to validate whether GEIGER achieves its goals and meets user requirements in operational environments. Validation according to this definition encapsulates more than measuring KPIs and meeting requirements. Validation entails subjecting the arguments and assumptions behind the GEIGER solution to rebuttals to see whether they are resilient enough to handle this burden.

To provide a rigorous basis for our validation work, we introduce a theoretical validation framework in this deliverable, which will be used throughout GEIGER's validation process. Our argument-based validation framework is based on earlier theoretical work in the field of educational measurement, specifically in formative assessment. Formative assessment in education offers an interesting parallel to the situation in GEIGER, as we are assessing MSEs and providing them with feedback to promote further learning and improvement.

This deliverable presents our practical validation planning, which is a direct result of our theoretical work. We discuss which activities will take place in each of three validation phases. The first phase of validation involves the GEIGER use case partners, which are spread out over the three GEIGER use case countries: Switzerland, The Netherlands, and Romania. The second phase involves a larger group of alpha users, and the third phase involves the largest group: beta users.

Although the second and third phase of validation are still ahead of us, we offer some first insights into the activities that these phases will be made up of. For the first phase of validation, we dive deeper into the validation activities that have already been conducted during the first six months of the validation work package WP4.

In discussing the details of validation, we note that validation activities often involve collaboration with other work packages. In performing technical experiments, we collaborate with and depend on the work done in WP2. Task T2.5 on optimization and hardening of the GEIGER solution is especially relevant for our validation work. Any of the WP4 work touching on aspects of the GEIGER education ecosystem involves intense collaboration and discussion with WP3, and the same is true of dissemination and WP5.

We additionally touch on the topic of data protection impact assessments (DPIAs) in this deliverable. Respecting the privacy and data protection concerns of participants in our WP4 experiments is vital, and we have invested a considerable amount of time to ensure we take the correct mitigating measures to reduce any risks in this domain. Together with work packages WP6 and WP7, we have devised a strategy to ensure that we adequately deal with the challenges of privacy and data protection in WP4.

We close this deliverable with an assessment of the progress made so far in validation. Although the groundwork for validation has been laid in the months preceding this deliverable, much is left to do in the months that remain in the project. We look forward to aiding the development of the GEIGER solution from a rigorous validation perspective in the final year of our project.

1 Introduction

GEIGER envisions a European landscape where Micro- and Small Enterprises (MSEs) are aware of current cybersecurity risks and resilient to any cyberattacks. By aiding MSEs in improving their knowledge regarding the topics of cybersecurity, privacy, and data protection, GEIGER empowers MSEs to independently improve their cybersecurity posture further. The ideal is an ecosystem where GEIGER serves as a platform for MSEs and their employees to help each other to continuously improve, creating a cyber resilient Europe.

A complex and innovative solution such as GEIGER relies on certain assumptions and inferences for its validity. By chaining our assumptions, strengthened by scientific and business research, we merge all GEIGER components into a single relevant solution for micro- and small enterprises (MSEs) throughout Europe.

The relevance of GEIGER's objectives is clear from the cyber threat landscape MSEs are currently faced with. Designing a relevant solution has also been the primary focus of the GEIGER project so far. In this delivera ble, and in the work done within the validation work package WP4 in general, we critically assess the steps we have taken so far. Our goal in validation is to address the companion of relevance in design science research: rigour.

GEIGER can be viewed as a socio-technical system. Socio-technical systems (STS) are characterised by intense interactions between their components, meaning the components should always be considered together, rather than separately. Figure 1 gives a brief overview of GEIGER's components and their interactions.

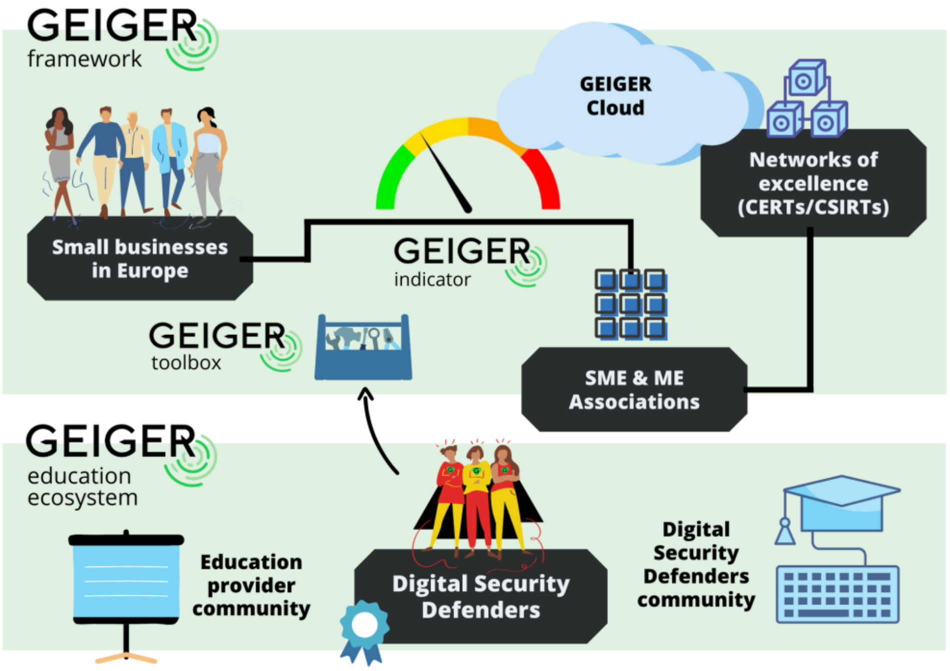

GEIGER framework
GEIGER Cloud
Networks of excellence (CERTs/CSIRTs)
Small businesses in Europe
GEIGER indicator
GEIGER toolbox
SME & ME Associations
GEIGER education ecosystem
Education provider community
Digital Security Defenders
Digital Security Defenders community

Figure 1: The GEIGER ecosystem.

The GEIGER components and their interactions is what makes the solution work. Concurrently, we must recognise that it is in the design of our components and in the instantiations of their interactions that our assumptions regarding the functioning of our solution can break down. In this deliverable, we unearth the key assumptions that have been made during the design process. We assess how critical the assumptions are and how large the 'leap of faith' that they make is.

Based on these assessments we determine where arguments are necessary to demonstrate the validity of our assumptions. This deliverable describes the arguments we deem necessary to address any potential worries, as well as the research methods that have been used, and are intended to be used, to provide the basis for our arguments.

In this sense this deliverable aligns with our overall objectives in the validation work package WP4:

- Validating GEIGER in three pilots (Switzerland, The Netherlands, and Romania) by applying the GEIGER solution in operational environments of the GEIGER use cases.
- Providing showcases for dissemination, standardisation, and policy.
- Performing the business experiments that are necessary for verifying the product-market alignment of the GEIGER solution, initiating the exploitation of project results.

WP4 work was initiated in month M13, meaning the work described in this deliverable spans the months M13-M18. During this time, we have achieved two main results:

1. The development of a validation framework based on an existing formative assessment validation framework. Our validation framework will serve as the theoretical basis for the validation activities in WP4, ensuring a rigorous approach.
2. The creation of an overarching validation planning for WP4. We distinguish three phases of validation: validation with use case partners, with alpha users, and with beta users. Each phase serves a unique purpose. Our modular design allows us to be flexible in adapting to any unforeseen delays within other work packages or the validation work package itself.

In the remainder of this deliverable, we first describe how the GEIGER objectives help to determine our objectives in validation (Section 2). We then present our novel validation framework in Section 3. Section 4 covers the validation planning that resulted from applying our validation framework. In Section 5, we dive deeper into specific validation activities that have taken place during the first validation phase.

Section 6 covers an important aspect during validation: data protection. We discuss how we assessed the potential risks regarding data collection for WP4 and which mitigation measures we employed to reduce these risks. Finally, in Sections 7 and 8, we look back at what we have achieved so far and offer our current take on the validity of the GEIGER solution.

2 Validation Overview

In the introduction, we briefly touched on the high-level objectives for WP4 as specified in the grant agreement. However, to guide validation efforts, we additionally need to specify validation objectives that are more closely tied to investigating whether we are meeting the goals GEIGER set itself. Therefore, we will first present an overview of the project vision and goals in Section 2.1, after which we define a more holistic definition of the validation objective in Section 2.2.

2.1 Project Overview

The GEIGER requirements deliverable (D1.1) laid out the GEIGER vision, challenges, and objectives in detail. We will briefly summarise these elements here to serve as an introduction for the rest of this deliverable.

The GEIGER solution is being built with the purpose of addressing a challenge for MSEs; the challenge of protecting MSEs against cyber-attacks and the potential harmful effects of negligent data protection. MSEs that fall victim to cyber-attacks risk interruptions of business, a loss of reputation, and potentially even bankruptcy. MSEs represent 98.9% of the enterprises in the European Economy[9], indicating the criticality of this challenge not only for the MSEs, but for Europe as a whole.

To address this challenge, GEIGER has set out the following goals that should be achieved if our solution is to be successful:

- The GEIGER solution should raise awareness of cyber threats that are personally relevant to MSE owners.
- The GEIGER solution should enable MSE owners to convert emotional coping into problem resolution.
- The GEIGER solution should support MSE owners in mitigating vulnerabilities at their company with suitable controls and a safeguarding security culture.

For further details on the GEIGER vision and objectives, please consult D1.1. In this deliverable, we will primarily focus on the implications of the GEIGER vision and objectives for the validation activities in WP4.

2.2 Validation Objectives

From the discussion in the previous section, we identified two sets of goals within the GEIGER project: the initial goals of GEIGER as measured by key performance indicators (KPIs) and newly discovered, more specific goals (or requirements) that resulted from the requirements elicitation work as described in deliverable D1.1 Requirements.

Both the overall GEIGER goals as measured by KPIs introduced in the grant agreement, and the GEIGER feature requirements resulting from the work of the Requirements work package WP1, should serve as input to our validation approach. This gives rise to a more holistic objective for GEIGER validation:
Validation Objective: *We aim to validate whether GEIGER achieves its goals and meets user requirements in operational environments.*

In layman's terms, this boils down to: does GEIGER do what it is supposed to do? The 'operational environment' qualifier is required since GEIGER aims to deliver a product achieving technology readiness level (TRL) 7: a system prototype demonstration in an operational environment. The intermediate TRL5 needed to achieve TRL7 corresponds to a technology validated in a relevant environment.

[9] See: European Commission (2018/2019): Annual Report on European SMEs 2018/2019. https://op.europa.eu/s/owB6

In the remainder of this section, we will first discuss exactly how KPIs and feature requirements are linked to validation (Sections 2.2.1 and 2.2.2). We will then also briefly discuss how the work of other GEIGER work packages can be seen in relation to our validation work (Section 2.2.3).

2.2.1 KPIs

KPIs make goals measurable. By linking one or more KPIs to a goal, we allow ourselves to measure progress regarding this goal. At some point, all KPIs related to a specific goal will be met and we are then able to judge whether the goal itself has been met. Note that this implies that meeting KPIs related to a goal is a minimum requirement to achieving the goal itself. Therefore, in the context of validation, measuring and meeting KPIs is necessary, not sufficient.

This points to the role KPIs play in GEIGER validation. We must ensure that our experiments and other data collection activities provide the data necessary to measure KPIs. The measurement results indicate relative progress towards meeting KPIs, which can serve as feedback to other work packages regarding the amount of work that remains to achieve our goals.

Therefore, KPIs help to guide our decision-making regarding employed research methods. They also help to communicate progress both within and outside the GEIGER consortium. Yet, using KPIs alone is insufficient in demonstrating validity. To understand why KPIs are necessary but not sufficient, let us consider an example in the GEIGER context.

One goal of the GEIGER solution is to make MSE owners aware of cyber threats relevant to their company. In the GEIGER Requirements Deliverable D1.1 we related three KPIs to this goal:

- KPI 1.2: Understanding of GEIGER Risk Indicator ≥ 4.0 on 5-point mean opinion score (MOS) scale.
- KPI I2.1.2.1: Perceived level of risk transparency ≥ 4.0 on 5-point mean opinion score (MOS) scale.
- KPI I2.1.2.3: Perceived level of risk explanation ≥ 4.0 on 5-point mean opinion score (MOS) scale.

Although these KPIs clearly relate to the goal, they can never give us insight into all intricacies involved with the goal. We may ask: Does the GEIGER indicator cover relevant threats to individual MSEs? Does understanding equate to a creation of awareness? Does a general understanding by an MSE that they are facing risks equate to knowledge about specific threats? Many more questions can be raised in this case.

The point of this example is not to discount the value of KPIs. KPIs are still central to making the GEIGER validation approach measurable. Nevertheless, KPIs are necessary, but not sufficient. This explains the requirement for a theoretical validation framework, as presented in Section 3. The argument-based approach we outline in Section 3 creates the opportunity to address the questions posed in the previous paragraph, while not simply being an open-ended mandate stating that all possible questions and worries must be addressed.

2.2.2 Feature Requirements

In the Requirements Deliverable D1.1 and the Training Plan Deliverable D3.1 we translated user needs into GEIGER feature requirements. We distinguished three types of features: cloud features, toolbox features, and educational features. Each relates to a specific component in the overall GEIGER ecosystem as presented in Figure 1. Together, these features serve the purpose of providing an optimal GEIGER customer journey.

By incorporating the GEIGER customer journey directly into our theoretical validation framework, we address many of the feature requirements automatically. Nevertheless, the features resulting from our requirements may at times be more granular than the elements of the GEIGER customer journey. Here, features can serve to guide the choice for exact implementations of validation activities.

An example is a feature resulting from a requirement from a specific use case country. Validating this feature may not be explicitly mentioned in our overall validation approach, but it is a necessary activity in the use case country in question. The implementation of a particular research activity in that use case country may then differ to other implementations to ensure the country-specific feature is appropriately validated.

Requirements and their related GEIGER features therefore serve two main purposes within validation:

1. Acting as a sanity check that GEIGER incorporates the necessary features in its solution.
2. Guiding the specification of certain validation activities where general validation activities do not address the validation needs related to a feature.

Although requirement features offer guidance as to what we can assess the validity of, it does not specify which questions we should ask or which indicators we should measure. Additionally, basing validation on feature requirements ignores the interrelation and interaction between features which is central in an STS like GEIGER.

To summarize, we have seen that KPIs and feature requirements help to guide validation and make it measurable. However, we have also seen that building a validation strategy based purely on KPIs and feature requirements is insufficient to meet our validation objective. It ignores key questions related to the content, constructs, and criteria used. It additionally does not consider the complex interactions within the GEIGER project. We will discuss how we intend to address these issues with our theoretical validation framework in Section 3.

2.2.3 Related Work Packages

The work done in WP4 will generally be relevant to the work packages involved in design, as our results provide feedback on the progress that has been made and the issues that remain. Similarly, the design research and work done in other work packages will at times serve as input for validation argumentation. We will briefly discuss the relation of WP4 to other work packages in this section.

As we noted in the previous section, WP1 on Requirements, Architecture, and Methodology provides input for validation in the form of feature requirements. Additionally, it specified the GEIGER architecture and aspects of the GEIGER indicator and GEIGER education methodologies. Nevertheless, WP1 work finished in M12, meaning the WP4 work which started in M13 has essentially no overlap with WP1. In this sense, WP1 provides much of the inspiration for what we should do in WP4 and why but does not collaboratively contribute to WP4.

WP2 focuses on the development of the technical GEIGER solution. Work for WP2 started in M04 and will finish at the end of the GEIGER project, in M30. Hence, there is a large overlap between WP2 and WP4 work. In WP4, we rely on the architectural and technical definitions of the GEIGER solution to inform decisions on how to evaluate which elements of the GEIGER application. Concurrently, validation activities in operational environments for WP4 will provide input for the design and development activities in WP2.

A specific task of WP2 worth mentioning is task T2.5 on Optimization and Security Hardening, which has the same runtime as WP4: M13-M30. Task T2.5 focuses on testing the integrated GEIGER solution, with the goal of optimizing the framework and improving its resilience. Some of the technical experiments performed in T2.5 can serve as direct input for validation argumentation, as it can highlight the validity of components and their interactions. Reliability is one dimension of validity where collaboration between T2.5 and WP4 will likely be necessary.

The WP3 work package relates to the GEIGER Educational Ecosystem involving security defenders and their education. WP3 runs throughout the GEIGER project. Given that the theoretical validation framework we use in WP4 is largely inspired by validation in educational measurement, it is not surprising that we expect intense collaboration between WP3 and WP4 throughout the WP4 runtime. From evaluations on the validity of the educational curriculum to monthly surveys following the progress of certified security defenders (CSDs), many of the validation tasks will be relevant to both WP3 and WP4. In Section 5.1.4 we dive deeper into the potential areas for collaboration, and the areas where WP3 and WP4 take slightly different approaches.

WP5 relates to Dissemination and Exploitation. Recall that one of the high-level objectives of WP4 is to provide showcases for dissemination, standardization, and policy. Another is to verify product-market alignment and initiate exploitation of project results. Clearly, WP5 is another work package where many synergies with WP4 are possible. In WP4 we are dependent on the dissemination success in WP5 for the statistical validity of many of our conclusions, since such conclusions generally depend on having sufficient samples to assess. WP5, in turn, relies on the results from WP4 experiments to aid in the dissemination of GEIGER achievements.

Finally, WP6 on Project Management and WP7 related to Ethics Requirements help to inform our choices regarding data management and the privacy impact assessments related to our data processing activities. We will continue to collaborate with these work packages to ensure we meet any data regulation requirements. Section 6 discusses the topic of data protection impact assessments (DPIAs) in more detail.

We have seen in this section how the background of the GEIGER project will inform the choices made in WP4. We have stated our holistic objective regarding validation concisely. We identified a need for an overarching theoretical validation framework to guide our validation activities and discussed how collaborations with other GEIGER work packages will help us to achieve our goals. Next, we will discuss our novel theoretical validation framework.

3 Theoretical Framework

Validation should not entail a loosely connected set of activities, but rather a concerted and guided effort to structurally arrive at a validity assessment. To guide our validation efforts in GEIGER, we searched the scientific body of knowledge to develop our own validation framework. The framework is based on an existing validation framework for formative assessment by Hopster-den Otter et al. (2019). We extend the framework to ensure it suits the socio-technical setting of the GEIGER project. In the following sections we describe the building blocks of our approach and the results it generated.

3.1 GEIGER Journey and Personas

The GEIGER solution can be seen as a socio-technical system, involving both human and technical elements that are tightly interconnected. The field of socio-technical systems (STS) recognises the inherent inseparability of humans and the technology they use, often within an organisational context (Cooper and Foster, 1971).

Before initiating validation in any STS project, two prerequisites must be satisfied:

1. The objective(s) of the project must be known.
2. A complete model of the STS must exist.

We have discussed the objectives of GEIGER at length in the previous section. We covered how progression regarding objectives can be measured using predefined KPIs. Validation will assess whether KPIs have been met, but also needs to go further than just providing evidence for meeting KPIs. Validation should also provide evidence that KPIs were met in the manner envisioned by the GEIGER project. To facilitate this process, a complete model of the GEIGER STS is required, to reason about the validity of the choices made and the artefacts produced.

In STS, one can argue about when a model of a system is 'complete.' In our case, we state that a model is complete when all internal and external actors, as well as the interactions between these actors, have been defined. This is in line with the focus of STS on interactions between actors within a given (external) environment (Davis, 2014; Paja et al., 2015). Business Process Model and Notation (BPMN) and the Unified Modelling Language (UML) are the two most widely used languages to model system design and processes, but they are notoriously complex due to their generic nature (Recker et al., 2009; Halvorsrud et al., 2016b).

Since GEIGER's overall vision is user-focused, an important requirement for any model of the GEIGER ecosystem used for validation is that it is also user-focused. This will be true for most projects that aim to improve the situation of users through a formative assessment procedure. Formative assessment is a term used in the education field to signal that assessment is used as an intermediate judgement to provide the learner with feedback to improve in the future. This distinguishes it from summative assessment, where the assessment constitutes a final decision (pass/fail).

Common examples of STS projects that implement formative assessment are e-learning and e-health applications, as well as cybersecurity risk assessment applications like GEIGER. In the remainder of this report, we will refer to STS projects that implement formative assessment as STS-FAs.

So, to validate an STS-FA, we require a modelling language that is not unnecessarily complex and is user-centric. A logical choice is to consider a modelling language that models the GEIGER customer journey, since this is inherently user-centric. BPMN and UML could potentially be used to model customer journeys, but it is not their intended purpose and, as stated before, their complexity is not ideal (Halvorsrud et al., 2016b). STS-ml is a language specifically aimed at socio-technical systems but cannot be used to model a customer journey since it is not temporal (Paja et al., 2015).

Since more common modelling languages are not suited to our requirements, we elected to use the Customer Journey Modelling Language (CJML) as our modelling language for validation (Halvorsrud et al., 2016a; Halvorsrud et al., 2016b). CJML models actors and their interactions, thereby satisfying our basic requirements for the modelling language used in validation. CJML has been shown to be suitable for modelling complex journeys in the healthcare domain (Halvorsrud et al., 2019), as well as user journeys in small- and medium-sized enterprise (MSE) cybersecurity solutions (Boletsis et al., 2021).

Especially this last fact provides a strong argument for using CJML in the GEIGER project, since GEIGER aims to help MSEs improve their cybersecurity posture. Table 1 outlines our motivations for the choice of CJML.

Table 1: Properties of several modelling languages that could be used to model the GEIGER solution.

Modelling Language	Complexity	User-Centric	Temporal
BPMN	High	Potentially	Yes
UML	High	Potentially	Yes
STS-ml	Medium	No	No
CJML	Low	Yes	Yes

Figure 2 shows a simplified version of the GEIGER journey modelled using CJML. We cover the complete GEIGER journey later in this section.

The GEIGER journey models the different types of use of the GEIGER solution. In STS, we refer to this as a 'role', distinguishing it from concrete (groups of) participants who are referred to as 'agents' (Paja et al., 2015). Whereas 'CSD' and 'User' are generic terms that refer to a specific role, 'John' and 'Laura' can be specific agents that perform such a role.

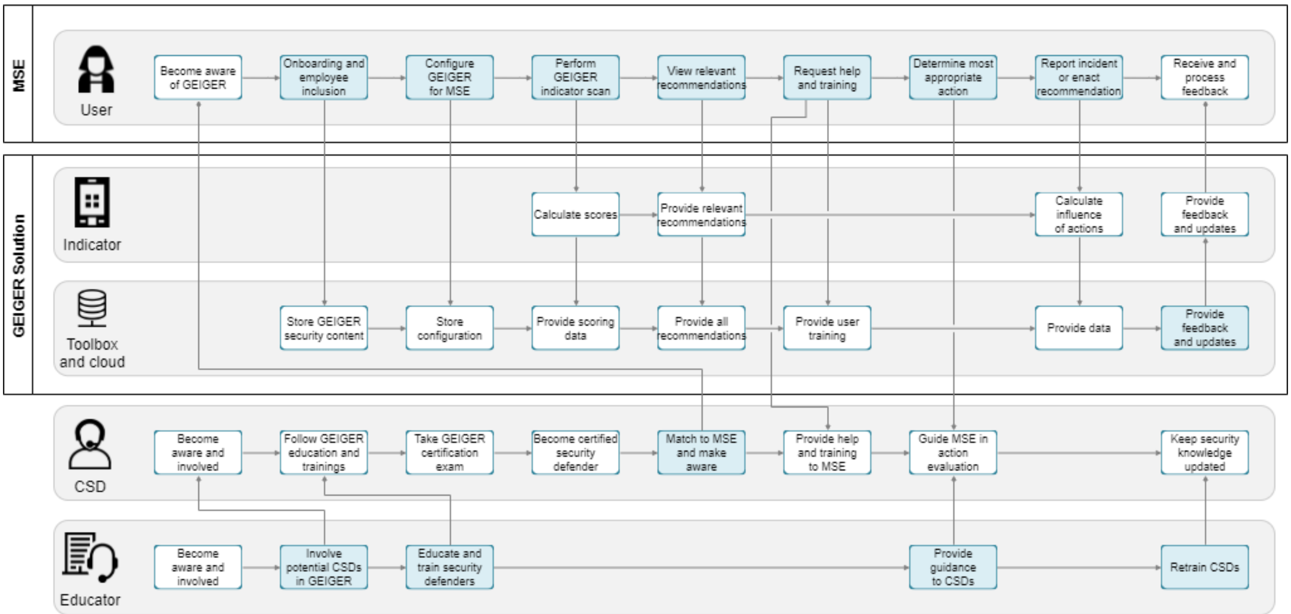
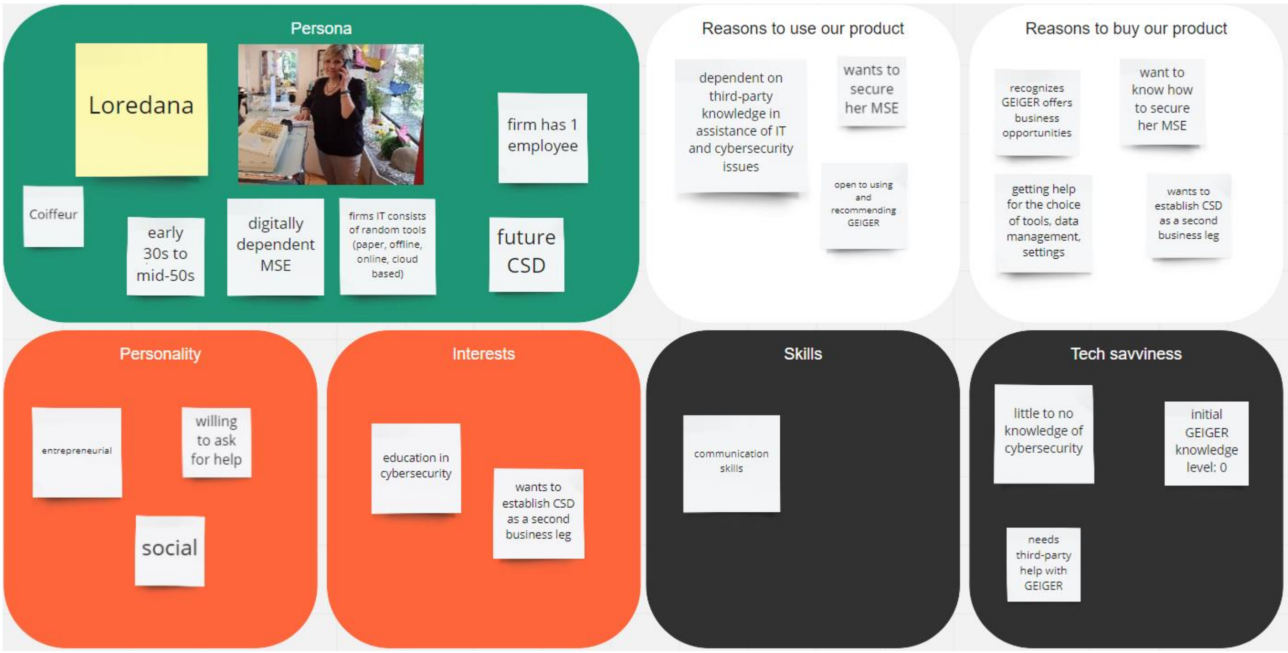

MSE
User
Become aware of GEIGER
Onboarding and employee inclusion
Configure GEIGER for MSE
Perform GEIGER indicator scan
View relevant recommendations
Request help and training
Determine most appropriate action
Report incident or enact recommendation
Receive and process feedback
GEIGER Solution
Indicator
Calculate scores
Provide relevant recommendations
Calculate influence of actions
Provide feedback and updates
Toolbox and cloud
Store GEIGER security content
Store configuration
Provide scoring data
Provide all recommendations
Provide user training
Provide data
Provide feedback and updates
CSD
Become aware and involved
Follow GEIGER education and trainings
Take GEIGER certification exam
Become certified security defender
Match to MSE and make aware
Provide help and training to MSE
Guide MSE in action evaluation
Keep security knowledge updated
Educator
Become aware and involved
Involve potential CSDs in GEIGER
Educate and train security defenders
Provide guidance to CSDs
Retrain CSDs

Figure 2: A simplified version of the GEIGER journey modelled using CJML.

During the validation process for GEIGER, we define personas that represent agents. One use or role can have several personas or agents. Each persona will have a slightly different GEIGER journey and they may all differ from the envisioned GEIGER journey for that role. CJML allows for this by distinguishing actual from planned journeys. The planned journey is how the journey is envisioned beforehand, often from a 'happy-flow' perspective. The actual journey occurs when a user instantiates the journey. In the actual journey, deviations are likely to occur compared to the planned journey, based either on errors in the solution or unexpected use of the solution.

By defining personas for GEIGER that cover a large set of our potential actors (including users, CSDs, educators, etc.), we hope to elicit the deviations from the planned journeys at an early stage, to be able to address any issues adequately during validation. Certain deviations may occur among all personas for a specific role, whereas at other times a deviation may be persona-specific.

In Figures 3, 4, and 5 we see examples of personas from the Swiss, Dutch, and Romanian use cases, respectively. We will dive deeper into the personas per use country in Section 5.2.



Persona
Loredana
firm has 1 employee
Coiffeur
early 30s to mid-50s
digitally dependent MSE
firms IT consists of random tools (paper, offline, online, cloud based)
future CSD

Reasons to use our product
dependent on third-party knowledge in assistance of IT and cybersecurity issues
wants to secure her MSE
open to using and recommending GEIGER

Reasons to buy our product
recognizes GEIGER offers business opportunities
want to know how to secure her MSE
getting help for the choice of tools, data management, settings
wants to establish CSD as a second business leg

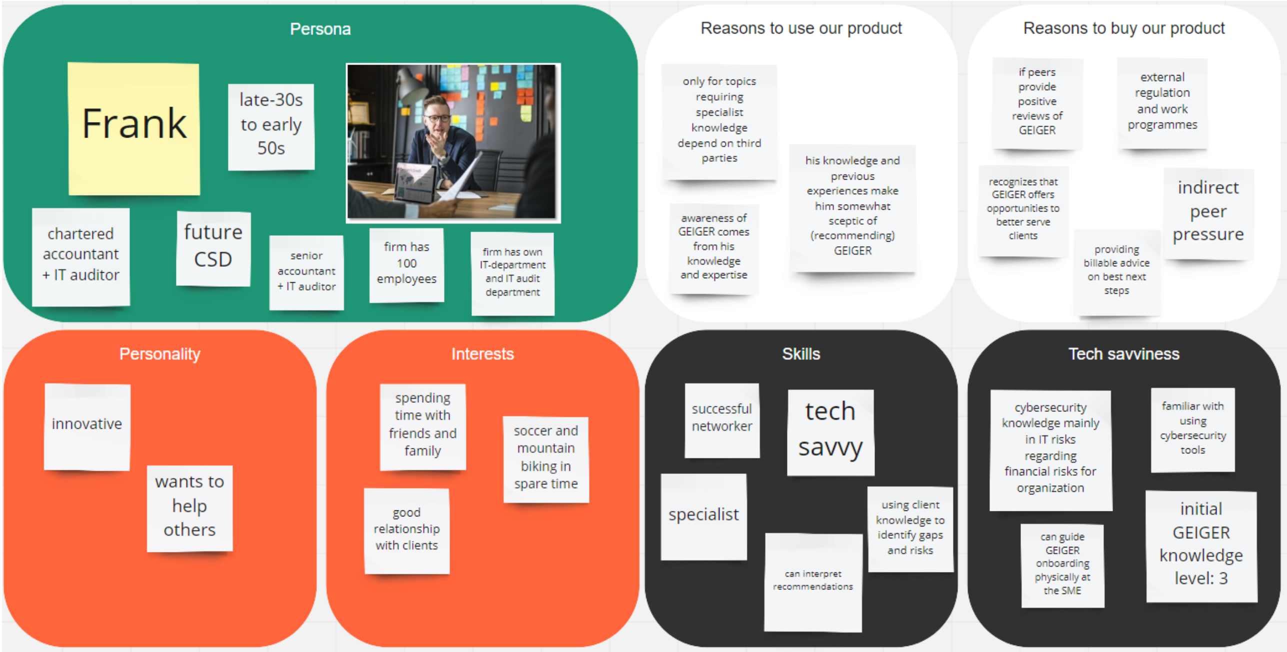
Personality
entrepreneurial
willing to ask for help
social

Interests
education in cybersecurity
wants to establish CSD as a second business leg

Skills
communication skills

Tech savviness
little to no knowledge of cybersecurity
initial GEIGER knowledge level: 0
needs third-party help with GEIGER

Figure 3: Loredana, one of the personas from the Swiss use case.



Persona
Frank
late-30s to early 50s
chartered accountant + IT auditor
future CSD
senior accountant + IT auditor
firm has 100 employees
firm has own IT-department and IT audit department

Reasons to use our product
only for topics requiring specialist knowledge depend on third parties
his knowledge and previous experiences make him somewhat sceptic of (recommending) GEIGER
awareness of GEIGER comes from his knowledge and expertise

Reasons to buy our product
if peers provide positive reviews of GEIGER
external regulation and work programmes
recognizes that GEIGER offers opportunities to better serve clients
indirect peer pressure
providing billable advice on best next steps

Personality
innovative
wants to help others

Interests
spending time with friends and family
soccer and mountain biking in spare time
good relationship with clients

Skills
successful networker
tech savvy
specialist
using client knowledge to identify gaps and risks
can interpret recommendations

Tech savviness
cybersecurity knowledge mainly in IT risks regarding financial risks for organization
familiar with using cybersecurity tools
can guide GEIGER onboarding physically at the SME
initial GEIGER knowledge level: 3

Figure 4: Frank, one of the personas from the Dutch use case.

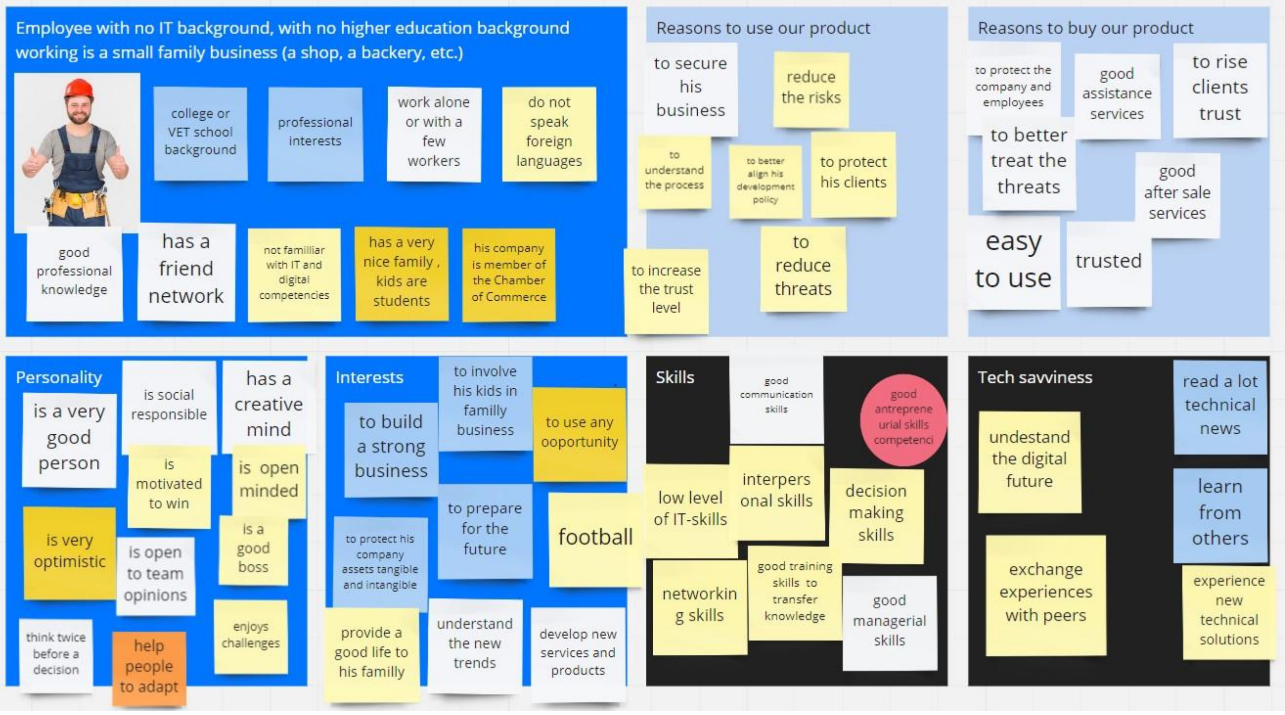

Employee with no IT background, with no higher education background working is a small family business (a shop, a backery, etc.)
college or VET school background
professional interests
work alone or with a few workers
do not speak foreign languages
good professional knowledge
has a friend network
not familliar with IT and digital competencies
has a very nice family , kids are students
his company is member of the Chamber of Commerce
Reasons to use our product
to secure his business
reduce the risks
to understand the process
to better align his development policy
to protect his clients
to increase the trust level
to reduce threats
Reasons to buy our product
to protect the company and employees
good assistance services
to rise clients trust
to better treat the threats
good after sale services
easy to use
trusted
Personality
is a very good person
is social responsible
has a creative mind
is motivated to win
is open minded
is very optimistic
is open to team opinions
is a good boss
think twice before a decision
help people to adapt
enjoys challenges
Interests
to build a strong business
to involve his kids in familly business
to use any ooportunity
to prepare for the future
football
to protect his company assets tangible and intangible
provide a good life to his familly
understand the new trends
develop new services and products
Skills
good communication skills
good antreprene urial skills competenci
low level of IT-skills
interpers onal skills
decision making skills
networkin g skills
good training skills to transfer knowledge
good managerial skills
Tech savviness
read a lot technical news
undestand the digital future
learn from others
exchange experiences with peers
experience new technical solutions

Figure 5: Florian, one of the personas from the Romanian use case.

3.2 Formative Assessment Validation

There are several approaches one could take to validate an STS-FA such as GEIGER. However, we should note that STS projects are inherently complex, making validation a challenging task. Therefore, we would ideally like to avoid "open-ended" (Kane, 2013) validation approaches where possible. Given the complexity of the GEIGER project, this is likely to lead to an impractical validation strategy.

This motivates the use of a validation framework in our GEIGER validation, to offer practical guidance to the validation task. Hopster-den Otter et al. (2019) formulated a validation framework for the validation of embedded formative assessment. Embedded, or "curriculum-embedded," formative assessment is the most formal type of formative assessment, which "consists of predefined tasks built into the school's educational program, that provide insights into students' current learning, and that is used to adapt teaching and learning to students' problem areas."

Embedded formative assessment is clearly linked to the GEIGER STS-FA, where a user performs "predefined tasks" that are captured in the GEIGER curriculum, to "provide insights into [their] learning" and cybersecurity posture, which are "used to adapt [countermeasure] and learning" suggestions to the user's "problem areas."

Figure 6 shows the interpretation and use argument (IUA) chain of the Hopster-den Otter et al. (2019) framework. We adapted the terminology to better suit our STS-FA context, as can be seen in Figure 7. The adaptation we made is to replace some of the terms that are commonly used in formative assessment with terms more common to socio-technical systems research. Specifically, we took inspiration from the terminology used in action design research (ADR), as covered in Sein et al. (2011).

The changes we made are relatively minor, but we believe they are necessary to ensure our approach aligns with the reality of the GEIGER project. In certain cases, such as when changing 'student learning' to 'learning,' the change originates from the fact that GEIGER does not operate in a traditional classroom setting. In other cases, such as with the change from 'decision' to 'outcome,' the reasoning is more subtle. This change was made to clearly communicate that the result of a GEIGER indicator scan is never final. GEIGER does not decide based on these results, but rather presents an outcome and looks to facilitate users in translating the outcome to their own situation (the rendition).

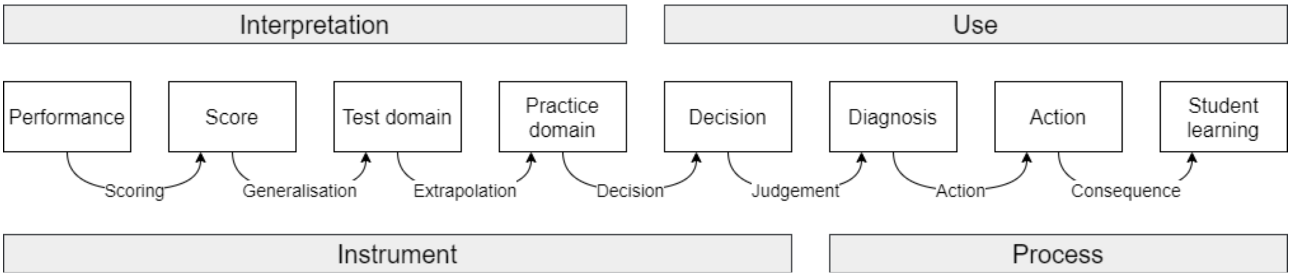


Figure 6: The Hopster-den Otter et al. (2019) IUA chain.

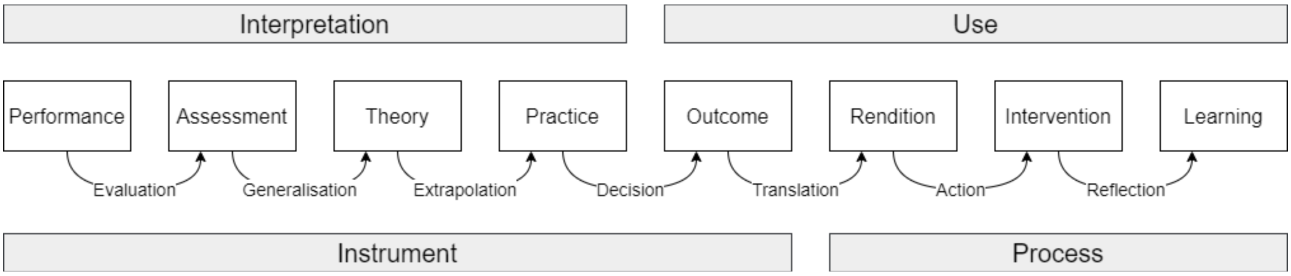


Figure 7: Our adapted version of the Hopster-den Otter et al. (2019) validation framework.

Figures 6 and 7 show how the steps of the argumentation chain are linked via inferences. In Figure 7, the 'Performance' and 'Assessment' arguments are linked by the 'Evaluation' inference, the 'Assessment' and 'Theory' arguments are linked by the 'Generalisation' inference, etc. We will cover specific details regarding each inference in Section 3.4. The important thing to realise for now is that validation using the framework requires two argumentation steps to be completed:

1. The interpretation and use argument (IUA): A complete specification of "the claims that are to be evaluated in the validation effort" (Kane, 2013).
2. The validity argument: "An evaluation of the proposed interpretation and use of test scores" (Kane, 2013).

Having just two steps to complete makes the approach sound simpler than it is. As we will see in next sections, there are still several actions required before we can begin with the validation of our STS-FA.

3.3 GEIGER Validation Framework

At the heart of the Hopster-den Otter et al. (2019) validation framework lies the concept of argument-based validation, first introduced by Kane (1992). In this section, we explain what argument-based validation entails and how argument-based validity differs from validity types such as criterion and construct validity. In Section 3.3.1, we cover the various validity types we focus on within the GEIGER project. Section 3.3.2 dives into the specifics of Toulmin arguments. Section 3.3.3 introduces the GEIGER arguments which together form the IUA for GEIGER. These arguments form the basis for the work in Section 3.4, where detailed guidelines for argument construction are presented.

3.3.1 Validity Types

Validity is a topic which has had many different interpretations over the years. It is also a concept which is interpreted differently in different research areas. Given that we employ a validation framework from educational measurement in the GEIGER project, we will focus on the concept of validity as it is defined in this field, while acknowledging the differences with the definitions employed in, for example, information systems and design science research.

Kane (1992) introduced argument-based validation with the intent to tackle some of the problems of "open-ended" (Kane, 2013) validation approaches that were in use at the time. Other common validity types discussed in the field of educational measurement are criterion validity, content validity, and construct validity. Table 2 outlines the definitions of these validity types, as well as the definition of argument-based validity.

Table 2: Common validity types in educational measurement, along with the argument-based approach to validity.

Validity Type/Approach	Definition	Source
Criterion validity	The extent to which test scores serving as an operationalisation of a construct correlate with, or predict, a theoretical representation of the construct (i.e., the criterion).	Cronbach and Meehl (1955)
Content validity	The extent to which "test items" are an appropriate "sample of a universe in which the investigator is interested."	Cronbach and Meehl (1955)
Construct validity	The extent to which a test measures what it claims to be measuring.	Cronbach and Meehl (1955)
Argument-based validation	**The extent to which the interpretation and use argument (IUA) for a test is "coherent and complete and all of its inferences and assumptions are highly plausible."**	**Kane (2013)**

It is important to note that all these types of validity are connected, and a validation approach of one type does not necessarily ignore other validation approaches. As Kane (2013) points out, for example, "criterion measures generally rely on the content model for their justification." Additionally, construct validity is at times seen as an overarching concept, and from the 1980s onwards was often "adopted as a general framework for validation" (Kane, 2013).

It is clear from Table 2 that the definitions originate from the psychological and educational measurement domain. However, criterion, content, and construct validity are now also common concepts in information systems research (Straub et al., 2004; Mingers and Standing, 2020). We observe that information systems research generally tends to define many different dimensions of validity, and to argue that an instrument is valid when all dimensions are sufficiently addressed (Straub, 1989). Besides the standard types of validity already mentioned, information systems research often considers internal validity, external validity, statistical (conclusion) validity, and reliability.

In Table 3, we list common definitions for the additional IS validity types mentioned. Straub (1989) groups content validity, construct validity, and reliability under the term "instrument validity". He additionally includes internal validity and statistical conclusion validity as vital elements to address to determine the validity of IS instruments. Straub (1989) does not explicitly address external validity but mentions that he does this "for the sake of brevity" rather than because he does not find it relevant.

Table 3: Additional validity types commonly encountered in IS research.

Validity Type	Definition	Source
Reliability	Answers the question: "Do measures show stability across the units of observation?"	Straub (1989)
Internal validity	Answers the question: "Are there alternative causal explanations for the observed data?"	Mingers and Standing (2020)
Statistical validity	Answers the question: Is our sampling approach sufficiently robust to rule out the possibility that results occurred by chance?	Straub (1989), Mingers and Standing (2020)

External validity	Answers the question: "To what extent can the findings be generalized to other populations and settings?"	Mingers and Standing (2020)

The definition of validity introduced by Cook et al. (2002, p. 34) is often used within the information systems field to represent their view of validity (italics in original text): "We use the term *validity* to refer to the approximate truth of an inference. When we say something is valid, we make a judgement about the extent to which relevant evidence supports that inference as being true or correct … Validity is a property of inferences. It is *not* a property of designs or methods." Especially this last remark is noteworthy, since it outlines the importance that inferences take within the validity definitions employed in information systems research.

The question we are left with is which approach to validation is best suited to validate our GEIGER STS-FA. You will have noted that argument-based validation was not mentioned as one of the validity dimensions mentioned in information systems. This is because argument-based validation "is an approach to validity rather than a type of validity" (Kane, 1992). The argument-based approach aligns well with the definition of validity used by Cook et al. (2002), and its generality allows us to capture the various types of validity considered important in information systems in our approach. Concurrently, the argument-based approach solves the issue often faced with other validity approaches that validation becomes an "open-ended, never-ending process," by providing a delimited framework (Kane, 2013).

Altogether we can conclude that the argument-based approach underlying the Hopster-den Otter et al. (2019) validation framework is suitable for the GEIGER project. Additionally, note that Straub's focus on instrument validity aligns well with this validation framework, which already creates a clear division between 'instrument' and 'process'. Figure 8 shows how we coupled the insights from Straub (1989), Straub et al. (2004), and Mingers and Standing (2020) with the existing framework of Hopster-den Otter et al. (2019) to form the GEIGER validation framework.

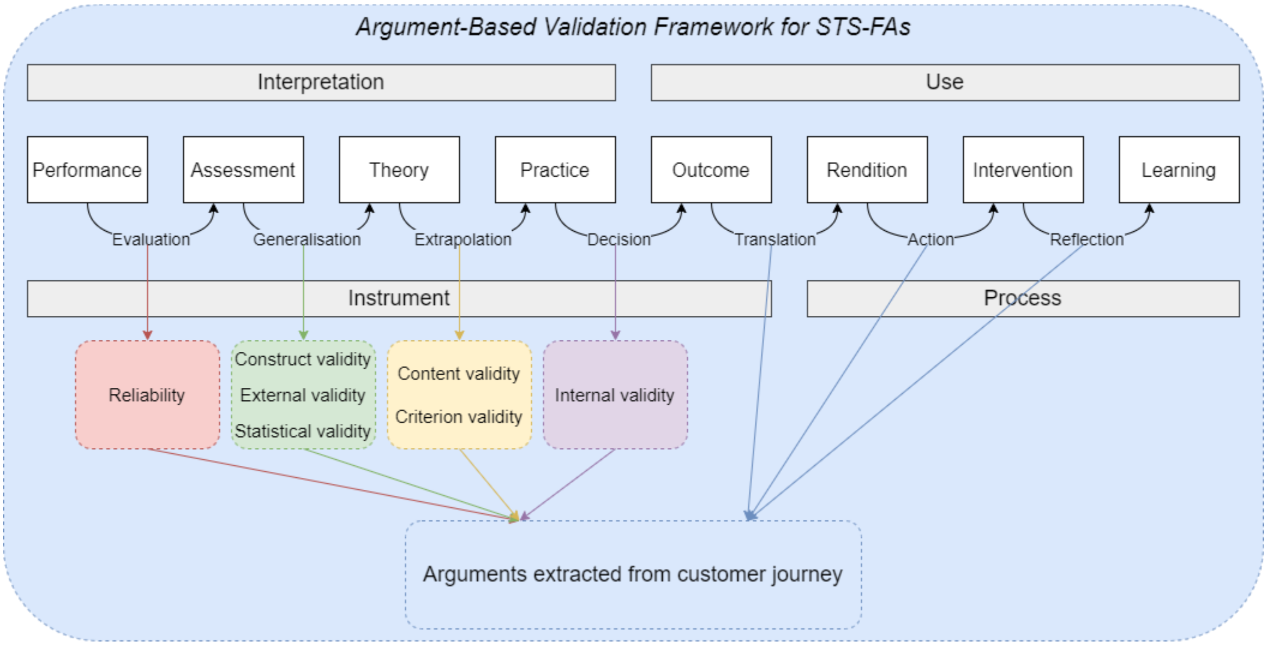


Figure 8: The GEIGER validation framework.

Figure 8 shows how our framework combines the validation knowledge bases of educational measurement and information systems into a single unified framework. The core idea for all IUA inferences is to extract our arguments in a structured manner by taking inspiration from our GEIGER journey. For the interpretation/instrument inferences we proceed in a different way than for the use/process inferences. The details of this procedure will be explained in Section 3.4. For now, it is sufficient to know that our process ensures we cover both the IUA inferences deemed vital in the Hopster-den Otter et al. (2019) framework, and ensures we address the core validity types of IS.

3.3.2 Toulmin Arguments

What constitutes an argument is not universally defined, and thus, when carrying out argument-based validation, one needs to choose the style of argumentation to use. In GEIGER we choose to use Toulmin arguments.

Toulmin arguments, developed by philosopher Stephen Toulmin (Toulmin, 2003), divide argumentation into several components: claim, data, warrant, qualifier, rebuttal, and backing. The claim, that which you want to prove or establish, is supported by data which are the "facts we appeal to as a foundation for the claim" (Toulmin, 2003). Warrants function as a "bridge" between the data and the claim, by showing "that the step from original data to the claim is a legitimate one" (Wools et al., 2010). If the claim is not expected to hold based on the data in all circumstances, and one wants to apply some nuance, a qualifier can be used to express this nuance. Rebuttals indicate "circumstances in which the general authority of the warrant would have to be set aside," and, lastly, backings are assurances in addition to the warrants, which are likely to be necessary in the presence of one or more rebuttals.

Figure 9 shows an example of what a completed Toulmin argument looks like, based on an argument outlined in Toulmin (2003). The exercise starts by trying to establish the claim that Petersen is not a Roman Catholic. We know that Petersen is a Swede, this is our data. Since a Swede can be taken to be almost certainly not a Roman Catholic (warrant), because the proportion of Roman Catholic Swedes is less than 2% (backing), we can say that almost certainly (qualifier) the claim holds true.

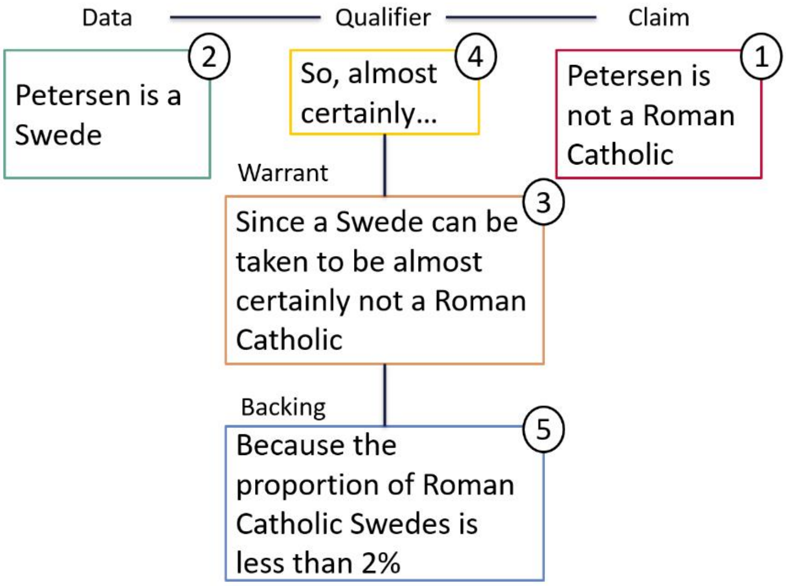

Data
Qualifier
Claim
2
Petersen is a Swede
4
So, almost certainly...
1
Petersen is not a Roman Catholic
Warrant
3
Since a Swede can be taken to be almost certainly not a Roman Catholic
Backing
5
Because the proportion of Roman Catholic Swedes is less than 2%

Figure 9: An example argument based on Toulmin (2003).

Using the Toulmin style of argumentation yields structured and clear arguments, which is why it is not surprising that Toulmin argumentation is a commonly employed argumentation style in argument-based validation (Simon, 2008; Wools et al., 2010). Our next step is to see how we can formulate Toulmin arguments for GEIGER, in each step of the IUA chain of the Hopster-den Otter et al. (2019) validation framework.

3.3.3 GEIGER Argumentation Chain

The IUA chain of the Hopster-den Otter et al. (2019) framework contains eight elements and seven inferences. In this section, we present the core claims, data, and qualifiers relating to the various steps of the chain from the GEIGER perspective. Section 3.4 will dive deeper into rebuttals, warrants, and backings, as well as the associated research methods we plan to use in validation.

Figure 10 shows the core elements of the GEIGER argument to infer an assessment from performance, using the evaluation inference. In this case, the 'data' that GEIGER provides a complete assessment of the cybersecurity risk of an employee and their devices, is clearly a more debatable statement than the example of the previous section (Figure 9), where the data was: "Petersen is a Swede." The implication of this observation is that even for this initial data we will need to provide sufficient warrants and backings to be able to use it to start our argumentation chain.

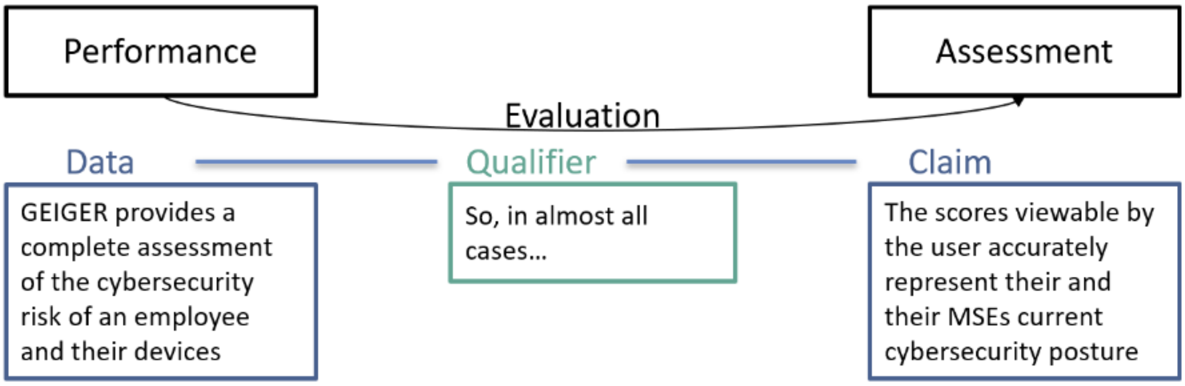

Performance
Assessment
Evaluation
Data
Qualifier
Claim
GEIGER provides a complete assessment of the cybersecurity risk of an employee and their devices
So, in almost all cases...
The scores viewable by the user accurately represent their and their MSEs current cybersecurity posture

Figure 10: Evaluation inference.

Once the performance claim has been established, we can use it as data in the argument of Figure 10. This process of using an earlier established claim as "starting-point for the next inference" (Wools et al., 2010), allows to iteratively move along the IUA chain. In the eventual appraisal of the IUA, an evaluation is made of how strong the links in the IUA chain are.

The evaluation inference builds on the assumption "that a set of scoring rules or algorithms provides insights into student learning strategies and mistakes" (Hopster-den Otter et al., 2019). Within the GEIGER context, this corresponds to the GEIGER indicator algorithm, which uses metrics resulting from the measurement of cybersecurity properties to provide an assessment of the user's cybersecurity posture. The feedback provided by the indicator in terms of GEIGER indicator scores, give the user an indication of where improvement is still required. Figure 8 showed how the evaluation inference can be linked to the concept of reliability.

The generalisation inference is depicted in Figure 11. Here, "we assume that the sample of tasks reflects the depth of student learning" (Hopster-den Otter et al., 2019), and "that the sample was large enough to control sampling errors" (Kane, 2013). Only if this is the case, will the GEIGER indicator scores viewed by the user provide an accurate estimate of the cybersecurity risk level faced by the user regarding the GEIGER topics of interest. In Figure 8, we saw that the generalisation inference is linked to construct validity, external validity, and statistical validity.

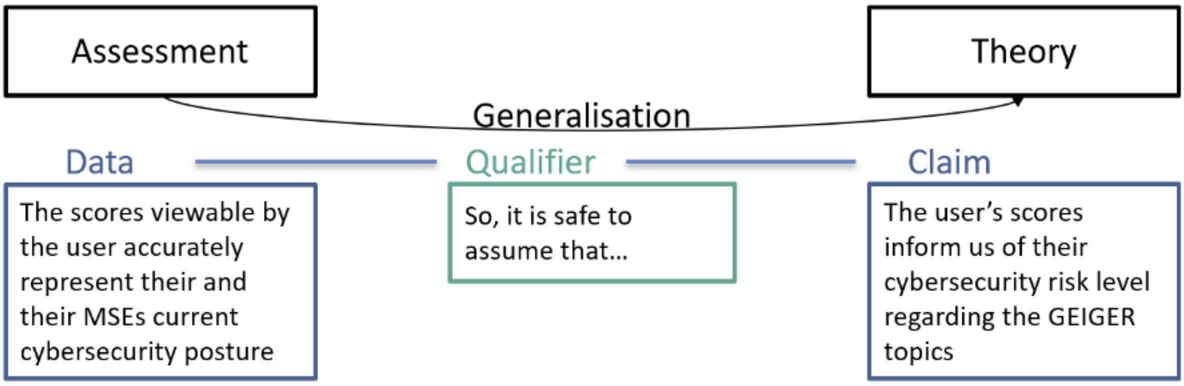

Assessment
Theory
Generalisation
Data
Qualifier
Claim
The scores viewable by the user accurately represent their and their MSEs current cybersecurity posture
So, it is safe to assume that...
The user's scores inform us of their cybersecurity risk level regarding the GEIGER topics

Figure 11: Generalisation inference.

Just establishing that a user's score reflects their cybersecurity risk level regarding the GEIGER topics is insufficient, since it may be the case that the GEIGER content does not reflect cybersecurity practice in the real world. The extrapolation inference, shown in Figure 12, addresses this point. In the extrapolation inference, we assume that theoretical tasks reflect practice. This inference relates to content- and criterion validity.

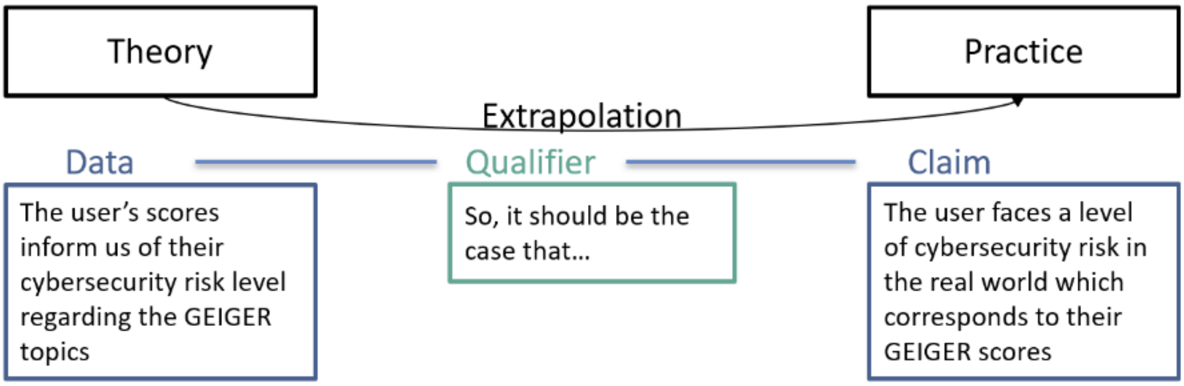

Theory
Practice
Extrapolation
Data
Qualifier
Claim
The user's scores inform us of their cybersecurity risk level regarding the GEIGER topics
So, it should be the case that...
The user faces a level of cybersecurity risk in the real world which corresponds to their GEIGER scores

Figure 12: Extrapolation inference.

In moving from the practice domain to the outcome (Figure 8), we move from the interpretation side of the argument chain to the use side. Hopster-den Otter et al. (2019) see formative assessment "as both an instrument and a process." The decision inference leading to the outcome is the final step of the instrument component of formative assessment.

The decision inference is based on "a decision rule that specifies how the decision will be made" (Hopster-den Otter et al., 2019). In GEIGER, a decision is made regarding the actions the user is suggested to take. In summative assessment processes, the decision would be final (e.g., pass/fail), which once more outlines why GEIGER applies formative assessment, rather than summative assessment, principles. The decision inference is linked to internal validity in the GEIGER validation framework.

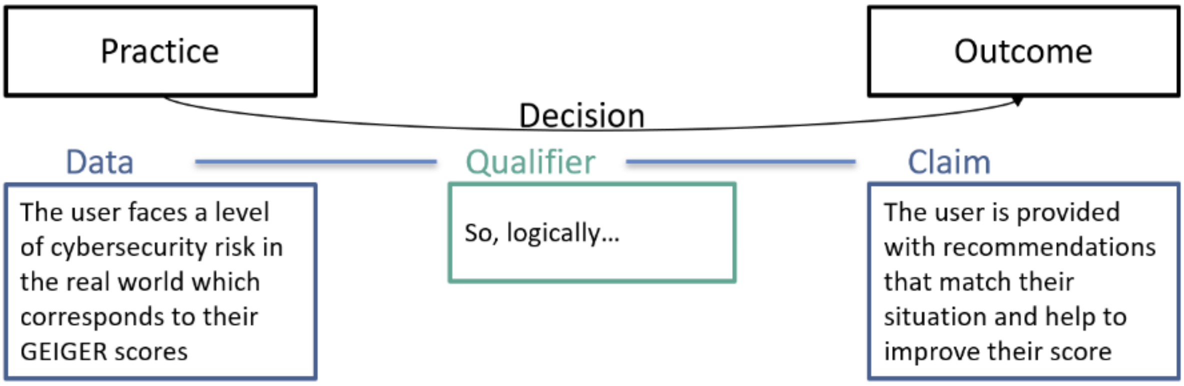

Practice
Outcome
Decision
Data
Qualifier
Claim
The user faces a level of cybersecurity risk in the real world which corresponds to their GEIGER scores
So, logically...
The user is provided with recommendations that match their situation and help to improve their score

Figure 13: Decision inference.

Figure 14 shows the translation inference, to move from an outcome to a rendition. In the rendition, we are evaluating use as part of the formative assessment process, meaning our focus has firmly shifted from the GEIGER instrument to the use of GEIGER.

In the translation inference "we assume that teachers and students are able to correctly understand the decision derived from the assessment instrument" (Hopster-den Otter et al., 2019). This is the first inference where the teacher is explicitly mentioned. For GEIGER, this implies the role of the CSD is vital for this argument, which is reflected in the definition of the claim. The user and the CSD should both understand and be able to interpret the provided recommendations. Additionally, the decisions should be linked and adapted to the user's context. The translation inference ties in well with MSE cybersecurity practice, since it is generally accepted that cybersecurity solutions for MSEs should adapt to user needs (Shojaifar et al., 2020).

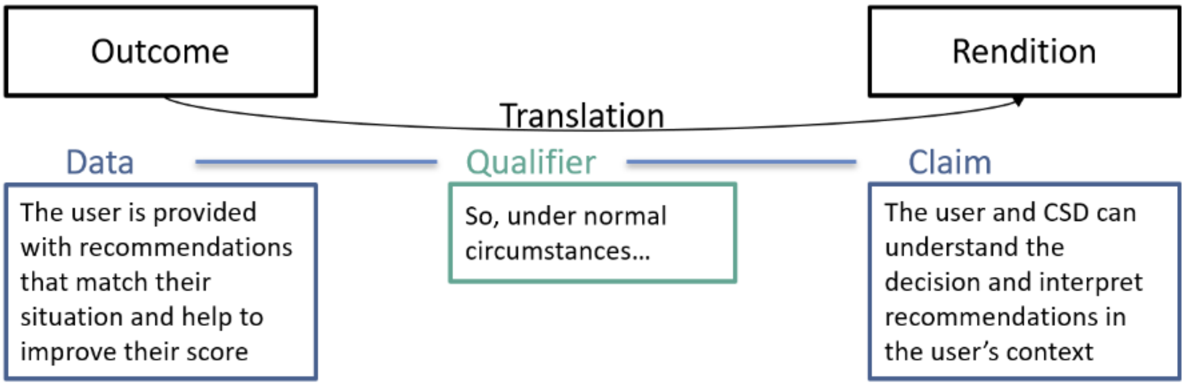

Outcome
Rendition
Translation
Data
Qualifier
Claim
The user is provided with recommendations that match their situation and help to improve their score
So, under normal circumstances…
The user and CSD can understand the decision and interpret recommendations in the user's context

Figure 14: Translation inference.

In the action inference of Figure 15, "we assume that the assessment information is tied to the curriculum and fits teachers' and students' knowledge base" (Hopster-den Otter et al., 2019). For GEIGER, this translates to the ability of the user, potentially aided by a CSD, to select appropriate and actionable recommendations. The user should also have a good sense of when they are done with a particular task, so that they can move on to a next task or consult the CSD for additional feedback.

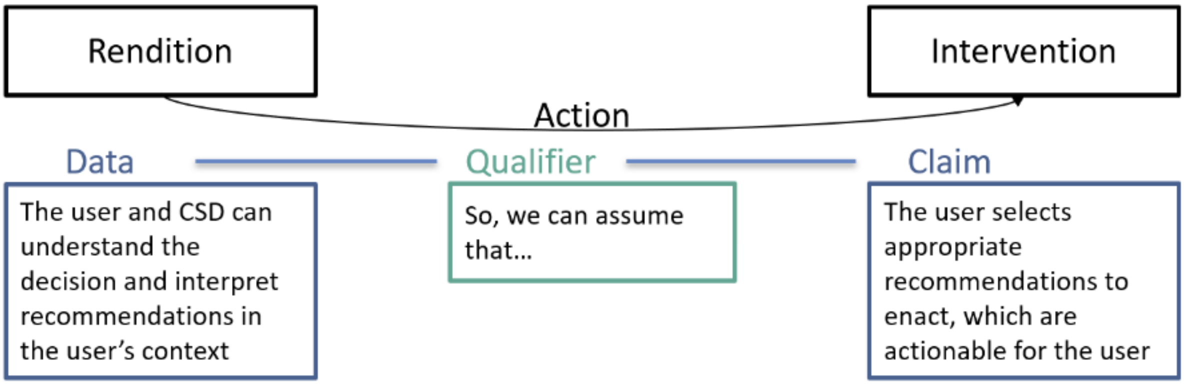

Rendition
Intervention
Action
Data
Qualifier
Claim
The user and CSD can understand the decision and interpret recommendations in the user's context
So, we can assume that…
The user selects appropriate recommendations to enact, which are actionable for the user

Figure 15: Action inference.

Lastly, the reflection inference is presented in Figure 16. In this inference "we assume that the approach to formative assessment results in student learning" (Hopster-den Otter et al., 2019), where in the GEIGER setting the term 'student' is removed. It is crucial to note that "this claim also assumes that the context is sufficiently supportive" (Hopster-den Otter et al., 2019). In the context of MSE cybersecurity, this means the user will likely need further support than purely receiving feedback from the GEIGER tool, meaning that once more the CSD has a vital role to play in this step of the GEIGER argument chain.

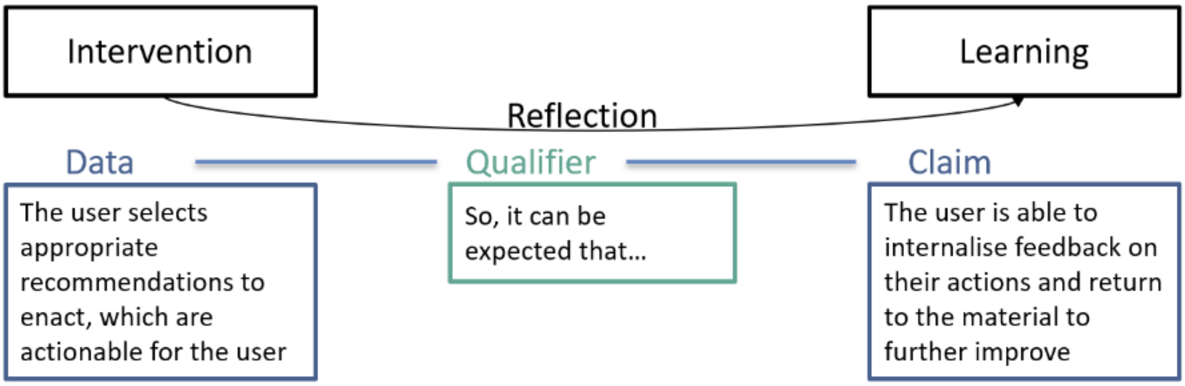

Intervention
Learning
Reflection
Data
Qualifier
Claim
The user selects appropriate recommendations to enact, which are actionable for the user
So, it can be expected that…
The user is able to internalise feedback on their actions and return to the material to further improve

Figure 16: Reflection inference.

When each step in the IUA chain has been sufficiently motivated, we move to the appraisal stage of the validation. Here, we argue for validity of GEIGER by examining "the coherence and completeness of the IUA and the plausibility of its inferences with respect to the purpose of [GEIGER's assessment]" (Hopster-den

Otter et al., 2019). Note that the appraisal does not yield a definitive conclusion that the GEIGER solution's assessment procedure is or is not valid, but rather a verdict on the plausibility of GEIGER's assessment being valid.

Although we have now detailed all data and claims forming the GEIGER IUA chain, we are still a long way from a complete, actionable validation approach for the GEIGER STS-FA. For this, we need to detail the warrants, rebuttals, and backings that originate from the GEIGER journey as modelled in CJML. This is what we aim to do in Section 3.4.

3.4 Argumentation Guidelines

The core argumentation elements discussed in Section 3.3.3 must be extended with appropriate warrants, rebuttals, and backings to arrive at a complete IUA chain for the GEIGER solution. The key issue is determining when these extensions are sufficient. Without demarcating our validation scope to some degree, we run the risk of getting stuck in the exact "open-ended, never-ending process" we intended to avoid (Kane, 2013).

The theory around applying Toulmin argumentation can help us in this regard. Figure 17, from Wools et al. (2010), shows how the Toulmin argument structure allows for an intuitive representation of a series of rebuttals and corresponding warrants and backings. By presenting argumentation in a visual manner, any reviewer evaluating a validation process is immediately given an impression of the level of detail achieved during validation.

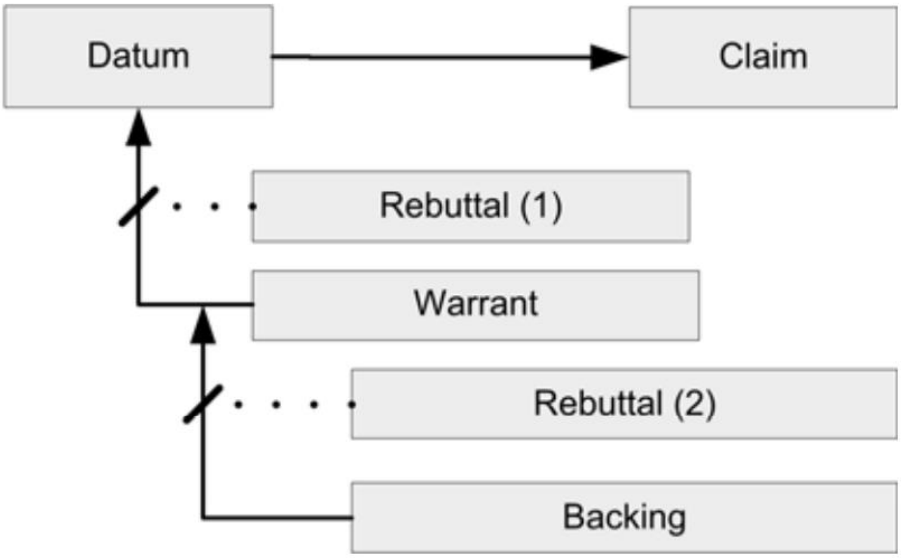


Figure 17: The Toulmin argument structure as presented in Wools et al. (2010).

Erduran et al. (2004) formalised this concept further, by introducing an analytical framework to assess argumentation quality. Table 4 shows the levels they defined, along with a description of what is expected of argumentation at each level. At the lowest level, Level 1, argumentation consists of no more than a claim versus a counterclaim. Progressing through the levels, we expect more of argumentation. By Level 5, we expect several rebuttals per argument, which are all adequately addressed.

Table 4: Table from Erduran et al. (2004), which covers an analytical framework used for assessing the quality of argumentation.

Level	Description
1	Argumentation consists of arguments that are a simple claim versus a counterclaim or a claim versus a claim.
2	Argumentation has arguments consisting of a claim versus a claim with either data, warrants, or backings but do not contain any rebuttals.

3	Argumentation has arguments with a series of claims or counterclaims with either data, warrants, or backings with the occasional weak rebuttal.
4	Argumentation shows arguments with a claim with a clearly identifiable rebuttal. Such an argument may have several claims and counterclaims.
5	Argumentation displays an extended argument with more than one rebuttal.

The framework of Erduran et al. (2004) allows us to formulate a minimal requirement to reach Level 5 of argumentation for the GEIGER solution:

For each argument of the GEIGER IUA chain, we consider argumentation to have achieved Level 5 when at least two rebuttals have been specified and have been adequately addressed through warrants and backings, for each role directly involved in the argument.

We consider a role to be directly involved in an argument when an action of that role within the GEIGER CJML journey is identified to be connected to the corresponding inference. Figure 18 illustrates this idea for a simplified version of the GEIGER journey.

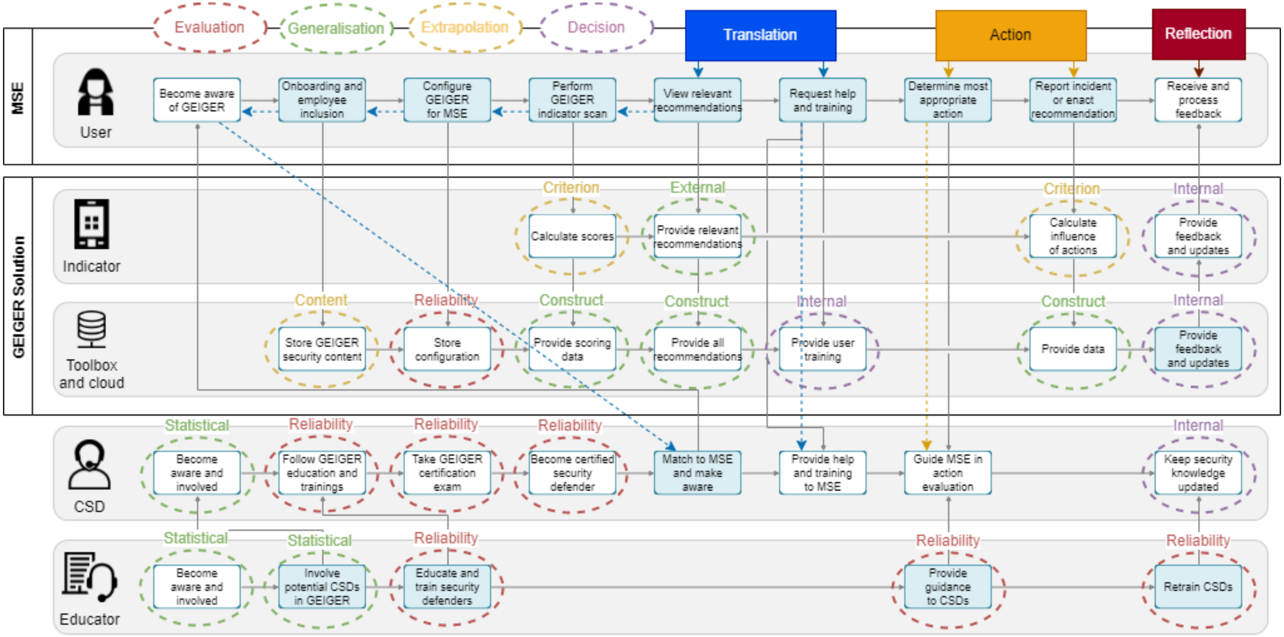

Evaluation
Generalisation
Extrapolation
Decision
Translation
Action
Reflection
MSE
User
Become aware of GEIGER
Onboarding and employee inclusion
Configure GEIGER for MSE
Perform GEIGER indicator scan
View relevant recommendations
Request help and training
Determine most appropriate action
Report incident or enact recommendation
Receive and process feedback
GEIGER Solution
Indicator
Criterion
Calculate scores
External
Provide relevant recommendations
Criterion
Calculate influence of actions
Internal
Provide feedback and updates
Toolbox and cloud
Content
Store GEIGER security content
Reliability
Store configuration
Construct
Provide scoring data
Construct
Provide all recommendations
Internal
Provide user training
Construct
Provide data
Internal
Provide feedback and updates
CSD
Statistical
Become aware and involved
Reliability
Follow GEIGER education and trainings
Reliability
Take GEIGER certification exam
Reliability
Become certified security defender
Match to MSE and make aware
Provide help and training to MSE
Guide MSE in action evaluation
Internal
Keep security knowledge updated
Educator
Statistical
Become aware and involved
Statistical
Involve potential CSDs in GEIGER
Reliability
Educate and train security defenders
Reliability
Provide guidance to CSDs
Reliability
Retrain CSDs

Figure 18: Simplified version of the GEIGER journey, with demarcations relating to argumentation.

To arrive at a completely demarcated journey, we follow the following steps:

1. Identify the entry points for use inferences.
2. Identify all use actions, meaning all actions by the user and all actions that are interactions between the user and another person.
3. For all remaining actions, determine the most appropriate validity type to associate with it and mark the action with this validity type.

In Step 1, we identify the entry points for use inferences. The three use inferences are translation, action, and reflection. In GEIGER, translation corresponds to the user internalising recommendations and translating them to their own situation, potentially with help from a CSD. The translation inference therefore corresponds to the 'View relevant recommendations' and 'Request help and training' actions. Action involves selecting recommendations to enact and enacting them. Potentially, action in GEIGER can also correspond to the reporting of an incident. Hence, the action inference is linked to the 'Determine most appropriate action' and 'Report incident or enact recommendation' actions. Lastly, the reflection inference involves the user processing feedback from the application to learn and further improve. This is evidently linked with the 'Receive and process feedback' action.

Step 2 involves identifying all use actions. We define a use action to be any action by the user participating in the formative assessment procedure, along with any interaction between that user and any other person(s) involved in GEIGER. This implies an action such as ‘Provide help and training to MSE’ is classified as a use action, since it involves a CSD helping a user. An action such as ‘Provide guidance to CSDs’ is not classified as a use action, since this involves an educator aiding a CSD, rather than a user.

Once we have completed Step 2, we now have all use actions clearly marked in our GEIGER journey. All remaining actions will be used to inform the arguments for the interpretation inferences, which we linked to the validity types associated with instrument validation. For each remaining action, we must evaluate which validity type is most appropriate to match with, and thus, which inference argumentation it will support. Figure 18 is the result achieved after completing the three steps.

Before diving into more details of the GEIGER IUA chain, we should cover our vision for where arguments can potentially be sourced from in our STS-FA setting. We want to provide guidance regarding suitable research methods. To assist this process, we consulted several papers which offer guidelines for validation in the context of information systems research and the argument-based approach. Table 5 presents these papers.

Table 5: Papers used as inspiration for determining relevant research methods.

Author and Year	Title
Straub et al. (2004)	Validation Guidelines for IS Positivist Research
Stockdale and Standing (2006)	An interpretive approach to evaluating information systems: A content, context, process framework
Pries-Heje et al. (2008)	Strategies for Design Science Research Evaluation
Fenz and Ekelhart (2011)	Verification, Validation, and Evaluation in Information Security Risk Management
Peffers et al. (2012)	Design Science Research Evaluation
Venable et al. (2012)	A Comprehensive Framework for Evaluation in Design Science Research
Wieringa (2014)	Design Science Methodology for Information Systems and Software Engineering
Cook et al. (2015)	A contemporary approach to validity arguments: a practical guide to Kane's framework
Mingers and Standing (2020)	A Framework for Validating Information Systems Research Based on a Pluralist Account of Truth and Correctness

Table 6 presents a sample of research methods to be found in some of the sources of Table 5 (Peffers et al., 2012; Venable et al., 2012; Wieringa, 2014; Mingers and Standing, 2020). However, the question remains where in our IUA chain we should be applying these methods. Cook et al. (2015) give some insights into this process, but purely for the education domain. To find more detailed guidance in our STS-FA setting, we look towards guidelines from design science research.

Table 6: Examples of research methods.

Research Method	Description	Source(s)
Action Research	Use of an artifact in a real-world situation as part of a research intervention, evaluating its effect on the real-world situation.	Peffers et al. (2012), Venable et al. (2012), Wieringa (2014)
Case Study	Application of an artifact to a real-world situation, evaluating its effect on the real-world situation.	Peffers et al. (2012) , Venable et al. (2012), Wieringa (2014)
Expert Evaluation	Assessment of an artifact by one or more experts.	Peffers et al. (2012), Wieringa (2014), Mingers and Standing (2020)
Field Experiment	Participants are divided into treatment and control groups. A treatment is applied to the	Venable et al. (2012), Wieringa (2014)

	treatment group, and we compare average outcomes to the control group.	
Illustrative Scenario	Application of an artifact to a synthetic or real-world situation aimed at illustrating suitability or utility of the artifact.	Peffers et al. (2012)
Logical Argument	An argument with face validity.	Peffers et al. (2012), Venable et al. (2012)
Prototype	Implementation of an artifact aimed at demonstrating the utility or suitability of the artifact.	Peffers et al. (2012)
Statistical Methods	Verifying convergence to envisioned constructs through statistical analysis of data.	Wieringa (2014), Mingers and Standing (2020)
Subject-Based Experiment	A test involving subjects to evaluate whether an assertion is true.	Peffers et al. (2012)
Survey	Questioning people to obtain data (in large numbers) to identify statistical regularities. Examples are paper questionnaires, web forms, and interviews.	Venable et al. (2012), Wieringa (2014)
Systematic Review	Reviewing literature in a reproducible manner to synthesise existing knowledge into useful insights.	Wieringa (2014), Mingers and Standing (2020)
Technical Experiment	A performance evaluation of an algorithm implementation using real-world data, synthetic data, or no data, designed to evaluate the technical performance, rather than its performance in relation to the real world.	Peffers et al. (2012)

To have some idea of when each research method of Table 6 could be used, we use the evaluation strategy and evaluation selection framework of Venable et al. (2012). In Venable et al. (2016), the framework was further extended to the Framework for Evaluation in Design Science Research (FEDS), but this framework is deemed unnecessarily complex for the task at hand.

We see in Table 7 that Venable et al. (2012) distinguish ex ante methods involving "an uninstantiated artifact," from ex post methods involving "an instantiated artifact." Additionally, the distinction between naturalistic and artificial methods is made. Naturalistic methods explore "the performance of a solution technology in its real environment," thereby embracing "all of the complexities of human practice in real organisations." This distinguishes it from artificial methods, whereby evaluation takes place in a controlled setting uncoupled from the real environment.

We can conclude from Table 7 that several research methods may be appropriate in every situation. Additionally, Table 6 and Table 7 are not intended as complete enumerations of research methods, but rather as extensive examples. If research methods that are not mentioned in these tables are appropriate in a specific scenario, they should certainly be employed.

Table 7: The GEIGER evaluation method selection framework, based on Venable et al. (2012).

Evaluation Method Selection Framework	Ex Ante	Ex Post
Naturalistic	• Action Research • Expert Evaluation • Subject-Based Experiment	• Action Research • Case Study • Expert Evaluation • Subject-Based Experiment • Survey

Artificial	• Illustrative Scenario • Logical Argument • Prototype • Statistical Methods • Systematic Review • Technical Experiment	• Field Experiment • Illustrative Scenario • Logical Argument • Statistical Methods • Technical Experiment

In Sections 3.4.1 and 3.4.2 we will formulate specific research method suggestions for each inference, based on the actions that were linked to that inference. Figure 19 shows the complete GEIGER journey, along with the actions and which validity types and inferences they were linked to.

Upon inspection we can see that all inferences are related to at least one action: evaluation (via reliability; 16 actions), generalisation (via construct, external, and statistical; 18), extrapolation (via content and criterion; 10), decision (via internal; 6), translation (13), action (3), and reflection (2). Especially the use inferences ‘action’ and ‘reflection’ have few related actions. Here we must take care to provide sufficient argumentation to meet the quality standards of Erduran et al. (2004) as outlined in Table 4.

For each inference we will also mention the relevant GEIGER KPIs for that inference. The KPIs are a key element in making the results of the GEIGER project quantifiable. Although KPIs are not always related to the validity of the GEIGER instrument, they do provide useful guidance for how to evaluate the GEIGER solution and tell us something about the relevance of the overall GEIGER solution. Combined with our rigorous approach to validation, the KPIs offer an avenue for addressing both relevance and rigour, two vital dimensions in design science research (Hevner et al., 2004).

As mentioned in Section 2.2.2, the GEIGER feature requirements can then be used at a later stage to determine whether adaptations to validation activities are needed to address specificities of particular use cases. As an example, a general validation activity may be adapted for each GEIGER use case country, based on the elicited feature requirements related to that activity.

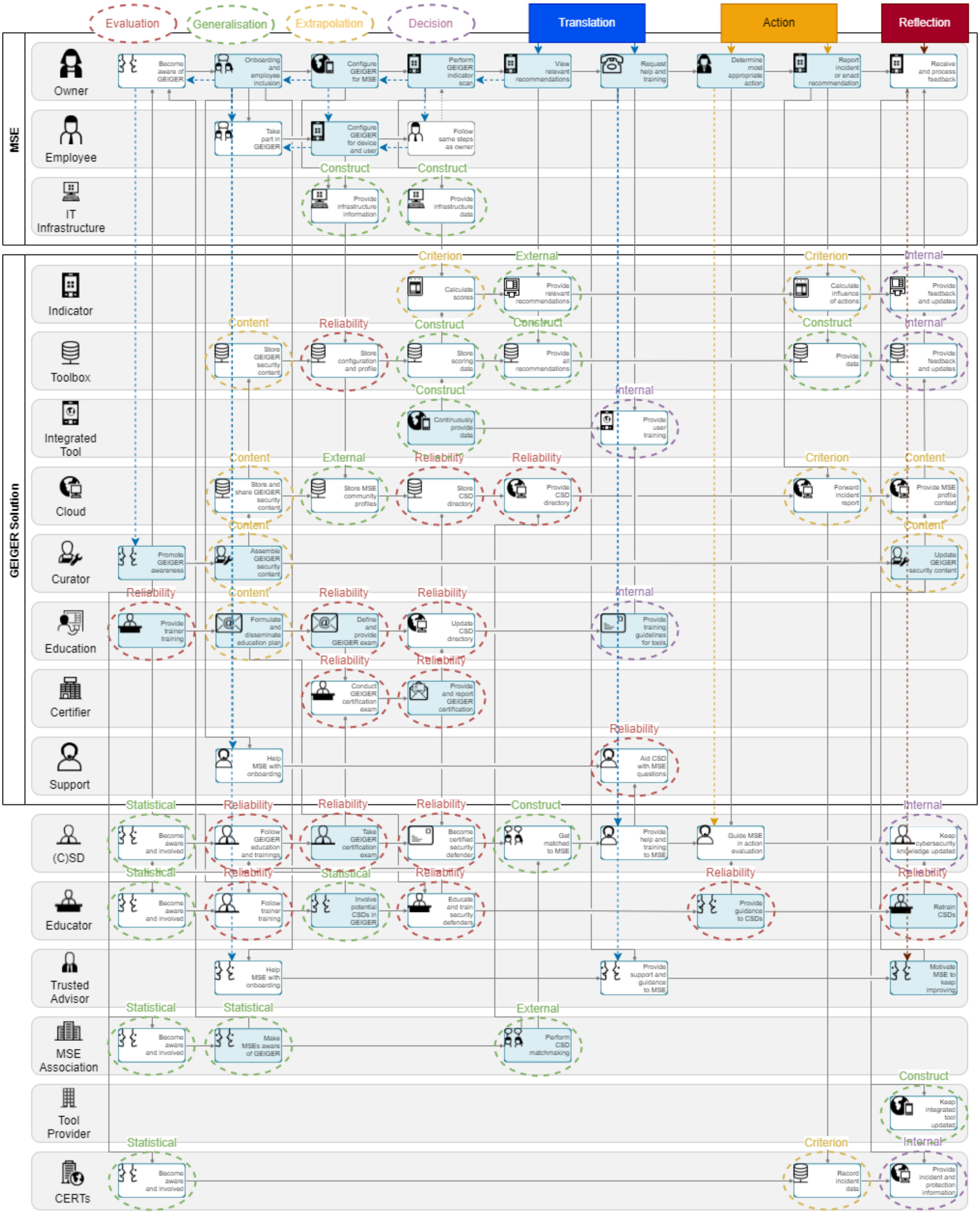

Evaluation
Generalisation
Extrapolation
Decision
Translation
Action
Reflection
MSE
Owner
Employee
IT Infrastructure
GEIGER Solution
Indicator
Toolbox
Integrated Tool
Cloud
Curator
Education
Certifier
Support
(C)SD
Educator
Trusted Advisor
MSE Association
Tool Provider
CERTs
Become aware of GEIGER
Onboarding and employee inclusion
Configure GEIGER for MSE
Perform GEIGER indicator scan
View relevant recommendations
Request help and training
Determine most appropriate action
Report incident or enact recommendation
Receive and process feedback
Take part in GEIGER
Configure GEIGER for device and user
Follow same steps as owner
Construct
Provide infrastructure information
Provide infrastructure data
Criterion
External
Internal
Calculate scores
Provide relevant recommendations
Calculate influence of actions
Provide feedback and updates
Content
Reliability
Store GEIGER security content
Store configuration and profile
Store scoring data
Provide all recommendations
Provide data
Continuously provide data
Provide user training
Store and share GEIGER security content
Store MSE community profiles
Store CSD directory
Provide CSD directory
Forward incident report
Provide MSE profile context
Promote GEIGER awareness
Assemble GEIGER security content
Update GEIGER security content
Provide trainer training
Formulate and disseminate education plan
Define and provide GEIGER exam
Update CSD directory
Provide training guidelines for tools
Conduct GEIGER certification exam
Provide and report GEIGER certification
Help MSE with onboarding
Aid CSD with MSE questions
Statistical
Become aware and involved
Take GEIGER education and trainings
Take GEIGER certification exam
Become certified security defender
Get matched to MSE
Provide help and training to MSE
Guide MSE in action evaluation
Keep cybersecurity knowledge updated
Follow trainer training
Involve potential CSDs in GEIGER
Educate and train security defenders
Provide guidance to CSDs
Retrain CSDs
Provide support and guidance to MSE
Motivate MSE to keep improving
Make MSEs aware of GEIGER
Perform CSD matchmaking
Keep integrated tool updated
Record incident data
Provide incident and protection information

Figure 19: The full GEIGER journey in CJML notation with each action coupled to a validity type or a use inference.

3.4.1 Interpretation/Instrument Validation

Figure 19 identified the actions in the GEIGER journey that can be linked to the first four inferences in the GEIGER IUA. These four inferences comprise the interpretation and/or instrument phase of our IUA chain.

For each inference, we will now proceed by identifying the actions corresponding to the inference and then determining which GEIGER KPIs can be linked to those actions. Additionally, we will investigate potential rebuttals related to the actions and suggest potential research methods that could be used to elicit warrants and backings to counter rebuttals. Altogether, the content of Section 3.4.1 and Section 3.4.2 will give the necessary input for the validation planning we perform in Section 4.

3.4.1.1 Evaluation Inference

The sole validity type associated with the evaluation inference is reliability. We can see from Figure 19 that there are a total of 16 actions related to reliability. The actions are listed in Table 8. We label each action using a capital 'A' along with identifiers for the inference and the validity type.

Table 8: Actions coupled to the evaluation inference.

Label	Validity Type	Role	Action
A.EV.RE.1	Reliability	Toolbox	Store configuration and profile
A.EV.RE.2	Reliability	Cloud	Store CSD directory
A.EV.RE.3	Reliability	Cloud	Provide CSD directory
A.EV.RE.4	Reliability	Education	Provide trainer training
A.EV.RE.5	Reliability	Education	Define and provide GEIGER exam
A.EV.RE.6	Reliability	Education	Update CSD directory
A.EV.RE.7	Reliability	Certifier	Conduct GEIGER certification exam
A.EV.RE.8	Reliability	Certifier	Provide and report GEIGER certification
A.EV.RE.9	Reliability	Support	Aid CSD with MSE questions
A.EV.RE.10	Reliability	(C)SD	Follow GEIGER education and trainings
A.EV.RE.11	Reliability	(C)SD	Take GEIGER certification exam
A.EV.RE.12	Reliability	(C)SD	Become certified security defender
A.EV.RE.13	Reliability	Educator	Follow trainer training
A.EV.RE.14	Reliability	Educator	Educate and train security defenders
A.EV.RE.15	Reliability	Educator	Provide guidance to CSDs
A.EV.RE.16	Reliability	Educator	Retrain CSDs

The actions of Table 8 have a strong relation to CSD training, examination, certification, and registration. This ties the evaluation inference to many of the KPIs related to the GEIGER Education ecosystem. CSDs that have followed proper GEIGER education will be able to ensure that the GEIGER experience for a varying group of MSEs is nonetheless the same. This is exactly what is referred to when we defined reliability as answering the question of whether measures show stability across units of observation. Both stability across MSEs, and stability of results within an MSE are promoted by a large CSD community supporting MSE users.

Table 9: KPIs related to evaluation inference actions.

Number	Description	Work Package
KPI 2.1	≥ 5 Capability areas addressed by training modules	WP3
KPI 2.2	≥ 2 Learning games	WP3
KPI 2.3	≥ 5 Cyber-range supported challenges	WP3
KPI 3.1	1 central GEIGER Cloud	WP2
KPI 5.1	3 "Certified Security Defenders" education approaches validated and demonstrated	WP3
Impact KPI I2.1.1.3	>50 MSEs will have benefitted from the Security Defender education in the Swiss pilot performed with apprentices by the school BBB	WP4

Impact KPI I2.1.1.4	>50 start-up MSEs will have benefitted from the Security Defender education in the Romanian pilot performed by the incubator/accelerator CLUJ IT CLUSTER	WP4
Impact KPI I2.1.1.5	>370 MSEs will have benefitted from advice by an accountant with Security Defender education in the Dutch pilot performed by the education provider ULEI	WP4
Impact KPI I2.1.1.6	>50 schools with vocational training for apprentices will intend to adopt the Security Defender education programme at the end of the project	WP4 (BBB)
Impact KPI I2.1.1.7	>50 incubators/accelerators will intend to adopt the Security Defender education programme at the end of the project	WP4 (CLUJ)
Impact KPI I2.1.1.8	>50 accountants education providers will intend to adopt the Security Defender education programme at the end of the project	WP4 (SRA)
Impact KPI I2.1.2.2	perceived level of decision support for risk reduction ≥ 4.0 on MOS scale	WP4
Impact KPI I2.1.2.9	4 contributions to standardisation work or MoUs with related initiatives for harmonising external GEIGER Framework interfaces and the security defenders education.	WP5
Impact KPI I2.1.4.4	≥50 education providers, incl. schools/universities, professional associations or unions, and incubators or accelerators for start-ups, will have confirmed their intent to offer the GEIGER education.	WP3
Impact KPI I2.1.5.2	≥200 educated Cyber Security Defenders	WP3
Impact KPI I2.1.5.3	≥100 certified Cyber Security Defenders	WP3

The education and training of CSDs is by no means a simple task. Therefore, several rebuttals can be defined regarding the various actions related to the evaluation inference. Besides the validation activities we perform in WP4 to address rebuttals, collaboration with WP3 is essential for this inference.

Table 10: Rebuttals to evaluation inference actions.

Label	Action	KPI	Rebuttal Description	Research Method(s)
R.EV.1	A.EV.RE.1	3.1	Toolbox storage does not store MSE configuration so that it is accessible for everyone in the MSE.	Illustrative Scenario, Technical Experiment
R.EV.2	A.EV.RE.2	3.1	The envisioned CSD directory does not function properly due to the combination of required personal information and privacy concerns.	Expert Evaluation, Survey
R.EV.3	A.EV.RE.3	3.1	The cloud environment is down when people try to access the CSD directory.	Illustrative Scenario, Statistical Methods
R.EV.4	A.EV.RE.4	2.1, 2.2, 2.3	The trainer training programme is not translated into all use case country languages.	Logical Argument
R.EV.5	A.EV.RE.5		The GEIGER exam is not available in all languages or not suited to all potential CSD target groups.	Expert Evaluation, Survey
R.EV.6	A.EV.RE.6		The education part of the consortium is not up to date on all CSDs.	Action Research, Case Study
R.EV.7	A.EV.RE.7		There is disagreement on what should be achieved during the GEIGER certification exam, leading to use case country differences.	Expert Evaluation, Survey

R.EV.8	A.EV.RE.8		The GEIGER certifying authority is insufficiently available to provide certification at any time.	Field Experiment, Statistical Methods
R.EV.9	A.EV.RE.9		The GEIGER support staff is not always available and does not always have the required knowledge to address CSD questions.	Expert Evaluation, Subject-Based Experiment
R.EV.10	A.EV.RE.10	5.1, I2.1.5.2	GEIGER education and training sessions are not accessible to all candidate CSDs.	Expert Evaluation, Subject-Based Experiment
R.EV.11	A.EV.RE.11		Certain potential CSDs feel their knowledge is already sufficient so they should be exempt from the GEIGER certification exam.	Survey
R.EV.12	A.EV.RE.12	I2.1.5.3	Potential CSDs would rather not receive the 'certified' title because of potential responsibilities associated with it.	Survey
R.EV.13	A.EV.RE.13		Educators do not have the time or motivation to follow GEIGER trainer trainings.	Action Research, Subject-Based Experiment
R.EV.14	A.EV.RE.14	I2.1.4.4	The education and training offered by the educator does not match the envisioned programme of the GEIGER consortium.	Case Study
R.EV.15	A.EV.RE.15		The educator has insufficient knowledge of the GEIGER environment to address CSD questions.	Subject-Based Experiment, Survey
R.EV.16	A.EV.RE.16	5.1, I2.1.2.9, I2.1.4.7	Educators do not have a programme in place to retrain CSDs when their knowledge should be updated.	Logical Argument

3.4.1.2 Generalization Inference

The generalization inference is linked to construct validity, external validity, and statistical validity. It is therefore not surprising that it covers a broad range of actions. We can see in Table 11 that construct validity in GEIGER relies heavily on the correct data being available at an MSE, as well as sufficient data being provided to an MSE.

External validity depends on our ability in GEIGER to keep providing relevant content to MSEs outside of our original use cases. We should still be able to provide relevant recommendations and promote feelings of relatedness at MSEs by matching them to CSDs and offering comparisons to scores of MSEs like them.

Statistical validity is largely dependent on whether we have enough MSEs in our sample to make sufficiently robust conclusions. Successful campaigns to raise awareness are therefore a necessity in this domain.

Table 11: Actions coupled to the generalisation inference.

Label	Validity Type	Role	Action
A.GE.CS.1	Construct	IT Infrastructure	Provide infrastructure information
A.GE.CS.2	Construct	IT Infrastructure	Provide infrastructure data
A.GE.CS.3	Construct	Toolbox	Store scoring data
A.GE.CS.4	Construct	Toolbox	Provide all recommendations
A.GE.CS.5	Construct	Toolbox	Provide data (e.g., current scores, implemented recommendations)

A.GE.CS.6	Construct	Integrated Tool	Continuously provide data (e.g., sensor values for indicator)
A.GE.CS.7	Construct	(C)SD	Get matched to MSE
A.GE.CS.8	Construct	Tool Provider	Keep integrated tool updated
A.GE.EX.1	External	Indicator	Provide relevant recommendations
A.GE.EX.2	External	Cloud	Store MSE community profiles
A.GE.EX.3	External	MSE Association	Perform CSD matchmaking
A.GE.ST.1	Statistical	(C)SD	Become aware and involved
A.GE.ST.2	Statistical	Educator	Become aware and involved
A.GE.ST.3	Statistical	Educator	Involve potential CSDs in GEIGER
A.GE.ST.4	Statistical	MSE Association	Become aware and involved
A.GE.ST.5	Statistical	MSE Association	Make MSEs aware of GEIGER
A.GE.ST.6	Statistical	CERTs	Become aware and involved

The diversity in actions for this inference can also be seen in the diversity in related KPIs and rebuttals. KPIs relate to the functioning of the indicator itself, the number of MSEs we will reach in the project, but also the level of tool usage among users.

Impact KPI ‘I2.1.3.2’ perhaps deserves some additional explanation, as the term ‘shield tools’ may have several meanings. In this context we intend ‘shield tools’ to refer to any tools in the GEIGER local toolbox or made available through the GEIGER cloud, that help to protect the MSE against cyber-attacks. We intend this to be interpreted in a broad sense, where even educational tools can be considered to protect the MSE, as they increase the awareness of employees, and thus the resilience of the MSE to attacks.

Table 12: KPIs related to generalisation inference actions.

Number	Description	Work Package
KPI 1.1	1 dynamic context-specific indicator of the current risk status (GEIGER Indicator)	WP2
KPI 1.3	Predicted attack intensity (e.g., affected devices) for a specific attack matches +/- 20% in more than 60% of all observed time spans retrospectively	WP4
KPI 5.2	GEIGER Framework will have been evaluated in ≥50 MSEs	WP4
KPI 5.3	Satisfaction by MSEs using the GEIGER Framework ≥ 4.0 on “Mean Opinion Score” scale ranging from 1 (bad) to 5 (excellent)	WP4
Impact KPI I2.1.1.1	>500’000 MSEs will be aware of the GEIGER Indicator as a dynamic risk monitoring instrument	WP5
Impact KPI I2.1.1.2	>50’000 MSEs will have tried the personalised GEIGER Indicator for their own specific MSE&ME by registering on GEIGER Solution	WP5
Impact KPI I2.1.2.4	≥1000 MSEs are connected to the GEIGER Cloud	WP2
Impact KPI I2.1.3.2	Shield tools are available and in use by the pilot MSEs for protecting against at least 80% of attacks recommended for protection by the participating CERTs/CSIRTs.	WP4
Impact KPI I2.1.4.1	≥1’000’000 impressions of the GEIGER Indicator as measured by number of impressions of media channels.	WP5
Impact KPI I2.1.4.2	≥100’000 small enterprises have a GEIGER account, allowing them to predict their risk with the personalised GEIGER Indicator and benefit from the GEIGER toolbox.	WP5
Impact KPI I2.1.4.3	≥20 MSE associations or chambers of commerce in ≥50% of the member states will have confirmed their intent to recommend the GEIGER Framework among their member enterprises.	WP5

Impact KPI I2.1.4.6	≥10 tools providers will have confirmed their intent to integrate their tools into the GEIGER toolbox.	WP2
Impact KPI I2.1.5.1	≥1’000 industry-diverse MSEs that know the GEIGER Indicator.	WP4

The technical experiments that need to be performed to address certain rebuttals clearly relate this generalization inference to task T2.5. The need to incorporate many MSE users to achieve statistical validity links this inference to WP5.

Table 13: Rebuttals to generalisation inference actions.

Label	Action	KPI	Rebuttal Description	Research Method(s)
R.GE.1	A.GE.CS.1		GEIGER has no connection to MSE IT infrastructures.	Prototype, Technical Experiment
R.GE.2	A.GE.CS.2		Data on the IT infrastructure of the MSE is not included in the GEIGER solution.	Logical Argument
R.GE.3	A.GE.CS.3		The GEIGER toolbox does not store data for a sufficiently long period.	Illustrative Scenario, Logical Argument
R.GE.4	A.GE.CS.4		The GEIGER toolbox does not have access to an updated list of all recommendations.	Illustrative Scenario, Technical Experiment
R.GE.5	A.GE.CS.5		The GEIGER toolbox does not adapt properly to all indications of implemented recommendations.	Statistical Methods, Technical Experiment
R.GE.6	A.GE.CS.6	I2.1.3.2, I2.1.4.6	Integrated tools do not update their data often enough.	Logical Argument, Technical Experiment
R.GE.7	A.GE.CS.7		CSDs refuse to be matched to MSEs due to anticipated travel times or other reasons.	Action Research, Survey
R.GE.8	A.GE.CS.8	I2.1.4.6, I2.1.2.9	Tool providers choose to not update their tools for GEIGER due to a limited number of GEIGER users.	Illustrative Scenario, Survey
R.GE.9	A.GE.EX.1		The indicator algorithm provides recommendations that do not pertain to the MSE situation, due to an incomplete view of the MSE profile.	Logical Argument, Prototype
R.GE.10	A.GE.EX.2	I2.1.2.4	MSE profiles cannot be stored in the GEIGER cloud.	Illustrative Scenario, Logical Argument
R.GE.11	A.GE.EX.3		MSE associations do not have enough knowledge about CSDs to perform appropriate matchmaking.	Case Study
R.GE.12	A.GE.ST.1	I2.1.5.2	Potential CSDs are not willing to become involved in the GEIGER education programme.	Action Research
R.GE.13	A.GE.ST.2	I2.1.1.6, I2.1.1.7, I2.1.1.8, I2.1.4.4	GEIGER cannot motivate sufficient educators to become involved in the project.	Logical Argument
R.GE.14	A.GE.ST.3		Educators are not willing to involve potential CSDs in GEIGER due to not believing in the solution.	Action Research, Case Study
R.GE.15	A.GE.ST.4	I2.1.4.3	MSE associations are unwilling to become involved.	Logical Argument, Survey
R.GE.16	A.GE.ST.5	I2.1.4.3	MSE associations have insufficient connections with MSEs to make them aware of GEIGER.	Case Study, Survey
R.GE.17	A.GE.ST.6	I2.1.4.5, I2.1.2.9	CERTs receive inadequate guidance in connecting to the GEIGER API.	Prototype, Survey

3.4.1.3 Extrapolation Inference

The extrapolation inference is linked to content validity and criterion validity. The GEIGER security content from both the indicator and the education perspective are relevant in this sense. Additionally, we can see that the criterion validity relates not only to the indicator score itself, but also to the idea of being able to predict any potential incidents.

Table 14: Actions coupled to the extrapolation inference.

Label	Validity Type	Role	Action
A.EX.CT.1	Content	Toolbox	Store GEIGER security content
A.EX.CT.2	Content	Cloud	Store and share GEIGER security content
A.EX.CT.3	Content	Cloud	Provide MSE profile context
A.EX.CT.4	Content	Curator	Assemble GEIGER security content
A.EX.CT.5	Content	Curator	Update GEIGER security content
A.EX.CT.6	Content	Education	Formulate and disseminate education plan
A.EX.CR.1	Criterion	Indicator	Calculate scores
A.EX.CR.2	Criterion	Indicator	Calculate influence of actions
A.EX.CR.3	Criterion	Cloud	Forward incident report
A.EX.CR.4	Criterion	CERTs	Record incident data

In a sense, incident occurrence is our gold standard criterion with which to measure the performance of the GEIGER indicator. We can see this directly in KPI 1.3, where we intend to use meeting the KPI as a proxy for a validity proof of the GEIGER indicator scoring mechanism. Other relevant KPIs are for example those measuring the capacity-building potential of GEIGER and KPIs related to incident sharing.

Table 15: KPIs related to extrapolation inference actions.

Number	Description	Work Package
KPI 1.1	1 dynamic context-specific indicator of the current risk status (GEIGER Indicator)	WP2
KPI 1.2	Understanding the GEIGER Indicator by CEOs of MSEs ≥ 4.0 on the Mean Opinion Score scale ranging from 1 (bad) to 5 (excellent)	WP4
KPI 1.3	Predicted attack intensity (e.g., affected devices) for a specific attack matches +/- 20% in more than 60% of all observed time spans retrospectively	WP4
KPI 2.1	≥ 5 Capability areas addressed by training modules	WP3
KPI 3.1	1 central GEIGER Cloud	WP2
KPI 3.2	4 open APIs allowing connectivity with MSEs, MSE associations, and CERTs/CSIRTs, and third-party tool and framework providers.	WP2
KPI 5.3	Satisfaction by MSEs using the GEIGER Framework ≥ 4.0 on “Mean Opinion Score” scale ranging from 1 (bad) to 5 (excellent)	WP4
Impact KPI I2.1.1.2	>50’000 MSEs will have tried the personalised GEIGER Indicator for their own specific MSE&ME by registering on GEIGER Solution	WP5
Impact KPI I2.1.1.9	GEIGER capacity-building assessed in surveys with >1’000 responses	WP4
Impact KPI I2.1.1.10	GEIGER capacity-building refined in >10 events targeting MSEs.	WP4

Impact KPI I2.1.2.1	perceived level transparency of risks ≥ 4.0 on MOS scale	WP4
Impact KPI I2.1.2.4	≥1000 MSEs are connected to the GEIGER Cloud	WP2
Impact KPI I2.1.2.5	≥3 CERTs/CSIRTs have access to the incident database	WP2
Impact KPI I2.1.2.6	≥3 data protection authorities have access to the incident database	WP2
Impact KPI I2.1.2.7	≥150 Security Defenders have access to the incident database	WP2
Impact KPI I2.1.2.8	1 open API with API access governance policies for querying incidents and submitting information	WP5
Impact KPI I2.1.2.9	4 contributions to standardisation work or MoUs with related initiatives for harmonising external GEIGER Framework interfaces and the security defenders education.	WP5
Impact KPI I2.1.3.2	Shield tools are available and in use by the pilot MSEs for protecting against at least 80% of attacks recommended for protection by the participating CERTs/CSIRTs.	WP4
Impact KPI I2.1.4.1	≥1'000'000 impressions of the GEIGER Indicator as measured by number of impressions of media channels.	WP5
Impact KPI I2.1.4.2	≥100'000 small enterprises have a GEIGER account, allowing them to predict their risk with the personalised GEIGER Indicator and benefit from the GEIGER toolbox.	WP5
Impact KPI I2.1.4.5	≥50% of the CERTs/CSIRTs in member states will have confirmed their intent to interoperate with the GEIGER Framework	WP5 (FHNW)
Impact KPI I2.1.5.1	≥1'000 industry-diverse MSEs that know the GEIGER Indicator.	WP4

Once more, this inference relies heavily on technical experiments, linking it to the optimization and hardening activities of T2.5. Additionally, we require expert evaluations of the GEIGER content, for example through the use of expert panels as described in Section 5.1.1.

Table 16: Rebuttals to extrapolation inference actions.

Label	Action	KPI	Rebuttal Description	Research Method(s)
R.EX.1	A.EX.CT.1		The toolbox is not properly integrated in the GEIGER solution.	Illustrative Scenario, Technical Experiment
R.EX.2	A.EX.CT.2	3.1, I2.1.2.4	The GEIGER cloud storage does not store and share data properly.	Illustrative Scenario, Technical Experiment
R.EX.3	A.EX.CT.3	3.1, 3.2	The GEIGER cloud is not able to provide MSE profile information on all devices within the MSE with GEIGER installed.	Illustrative Scenario, Technical Experiment
R.EX.4	A.EX.CT.4	1.1	The assembled security content is incorrect/inadequate for the MSE setting.	Expert Evaluation, Systematic Review
R.EX.5	A.EX.CT.5		There is not enough capacity at the GEIGER curator role to update all GEIGER security content regularly.	Logical Argument
R.EX.6	A.EX.CT.6	5.1	The education plan is not available in all languages (on time).	Logical Argument
R.EX.7	A.EX.CR.1	1.1	The GEIGER indicator algorithm does not have enough data to calculate all individual scores.	Prototype, Technical Experiment
R.EX.8	A.EX.CR.2		Implemented recommendations are not communicated properly to the indicator algorithm in all cases.	Statistical Methods, Technical Experiment

R.EX.9	A.EX.CR.3	I2.1.2.4	MSEs do not have a connection to the GEIGER cloud at all, or do not have a connection when they try to report an incident.	Statistical Methods, Technical Experiment
R.EX.10	A.EX.CR.4	3.2, I2.1.2.5, I2.1.2.6, I2.1.2.8	The GEIGER API connection with CERTs is insufficiently secure and reliable to be used for incident recording.	Illustrative Scenario, Technical Experiment

3.4.1.4 Decision Inference

The decision inference is linked to internal validity. Having alternative causal explanations for the results obtained becomes increasingly less likely as a user continues to use GEIGER and is assisted by a CSD who is constantly updating their knowledge based on the latest incident and protection information. This idea of a decreasing likelihood of alternative causal explanations over time, explains the actions contained in Table 17.

Table 17: Actions coupled to the decision inference.

Label	Validity Type	Role	Action
A.DE.IN.1	Internal	Indicator	Provide feedback and updates
A.DE.IN.2	Internal	Toolbox	Provide feedback and updates
A.DE.IN.3	Internal	Integrated Tool	Provide user training
A.DE.IN.4	Internal	Education	Provide training guidelines for tools
A.DE.IN.5	Internal	(C)SD	Keep cybersecurity knowledge updated
A.DE.IN.6	Internal	CERTs	Provide incident and protection information

The link to keeping knowledge updated and processing feedback can also be clearly seen in the KPIs related to the decision inference. Assuming everything is kept up-to-date, and this is made measurable by the below KPIs, we can convincingly argue for the claim associated to the decision inference.

Table 18: KPIs related to decision inference actions.

Number	Description	Work Package
KPI 1.1	1 dynamic context-specific indicator of the current risk status (GEIGER Indicator)	WP2
KPI 2.1	≥ 5 Capability areas addressed by training modules	WP3
KPI 2.2	≥ 2 Learning games	WP3
KPI 2.3	≥ 5 Cyber-range supported challenges	WP3
KPI 3.2	4 open APIs allowing connectivity with MSEs, MSE associations, and CERTs/CSIRTs, and third-party tool and framework providers.	WP2
KPI 5.1	3 “Certified Security Defenders” education approaches validated and demonstrated	WP3
KPI 5.3	Satisfaction by MSEs using the GEIGER Framework ≥ 4.0 on “Mean Opinion Score” scale ranging from 1 (bad) to 5 (excellent)	WP4
Impact KPI I2.1.1.3	>50 MSEs will have benefitted from the Security Defender education in the Swiss pilot performed with apprentices by the school BBB	WP4
Impact KPI I2.1.1.4	>50 start-up MSEs will have benefitted from the Security Defender education in the Romanian pilot performed by the incubator/accelerator CLUJ IT CLUSTER	WP4

Impact KPI I2.1.1.5	>370 MSEs will have benefitted from advice by an accountant with Security Defender education in the Dutch pilot performed by the education provider ULEI	WP4
Impact KPI I2.1.2.1	perceived level transparency of risks ≥ 4.0 on MOS scale	WP4
Impact KPI I2.1.2.2	perceived level of decision support for risk reduction ≥ 4.0 on MOS scale	WP4
Impact KPI I2.1.2.3	perceived level of risk explanation ≥ 4.0 on MOS scale	WP4
Impact KPI I2.1.2.5	≥3 CERTs/CSIRTs have access to the incident database	WP2
Impact KPI I2.1.2.8	1 open API with API access governance policies for querying incidents and submitting information	WP5
Impact KPI I2.1.3.1	at least 80% of basic recommendations for human error prevention are adopted by the pilot MSEs	WP4
Impact KPI I2.1.3.2	Shield tools are available and in use by the pilot MSEs for protecting against at least 80% of attacks recommended for protection by the participating CERTs/CSIRTs.	WP4
Impact KPI I2.1.4.5	≥50% of the CERTs/CSIRTs in member states will have confirmed their intent to interoperate with the GEIGER Framework	WP5 (FHNW)
Impact KPI I2.1.4.6	≥10 tools providers will have confirmed their intent to integrate their tools into the GEIGER toolbox.	WP2
Impact KPI I2.1.4.7	≥2 contributions to standardisation	WP5
Impact KPI I2.1.5.2	≥200 educated Cyber Security Defenders	WP3
Impact KPI I2.1.5.3	≥100 certified Cyber Security Defenders	WP3

The research methods to address the rebuttals for this inference need to consider the temporal nature of the rebuttals. This means the proposed surveys need to be performed over longer time periods, to determine whether it is indeed true that the GEIGER solution does not satisfy some of the updating requirements.

Table 19: Rebuttals to decision inference actions.

Label	Action	KPI	Rebuttal Description	Research Method(s)
R.DE.1	A.DE.IN.1		The indicator algorithm is not updated on the user's device due to a technical error.	Illustrative Scenario, Logical Argument, Technical Experiment
R.DE.2	A.DE.IN.2		The local GEIGER toolbox of the user is not updated due to no connection to the GEIGER cloud, either by choice or due to an error.	Action Research, Survey
R.DE.3	A.DE.IN.3	2.1, 2.2, 2.3	Integrated education tools are deemed inaccessible to users due to financial obstructions.	Survey
R.DE.4	A.DE.IN.4	2.1, 2.2, 2.3	The education ecosystem of GEIGER does not have an infrastructure in place to adequately communicate guidelines to tools.	Action Research, Case Study
R.DE.5	A.DE.IN.5	5.1, I2.1.2.2	CSDs change their priorities and do not keep their cybersecurity knowledge updated over time.	Survey
R.DE.6	A.DE.IN.6	3.2	The data provided by CERTs is not applicable to (some of) the MSEs situation.	Expert Evaluation, Survey

3.4.2 Use/Process Validation

In the use phase of the GEIGER IUA, argumentation is structured slightly differently since we do not work with explicit validity types. Rather, we see which actions of the GEIGER user journey by the user (the MSE owner) can be linked directly to an inference. From there, we determine the connected use actions. These can either be actions by the user earlier in the GEIGER journey that have not yet been addressed, or interactions with the user by other roles involving human actors.

To keep structurally the same labels, we replace the indicator for validity type with the first two letters of the directly connected use action.

3.4.2.1 Translation Inference

The two user actions connected to the translation inference are ‘View relevant recommendations’ (VI) and ‘Request help and training’ (RE). We additionally include all earlier user actions that have not yet been validated (and thus constitute a temporal dependence), and any interactions with the user and other persons related to included actions. This results in the inclusion of the actions in Table 20.

Table 20: Actions coupled to the translation inference.

Label	Role	Action
A.TR.VI.1	Owner	View relevant recommendations
A.TR.VI.2	Owner	Perform GEIGER indicator scan
A.TR.VI.3	Employee	Follow same steps as owner
A.TR.VI.4	Owner	Configure GEIGER for MSE
A.TR.VI.5	Employee	Configure GEIGER for device and user
A.TR.VI.6	Owner	Onboarding and employee inclusion
A.TR.VI.7	Employee	Take part in GEIGER
A.TR.VI.8	Support	Help MSE with onboarding
A.TR.VI.9	Trusted Advisor	Help MSE with onboarding
A.TR.VI.10	Owner	Become aware of GEIGER
A.TR.VI.11	Curator	Promote GEIGER awareness
A.TR.RE.1	Owner	Request help and training
A.TR.RE.2	(C)SD	Provide help and training to MSE
A.TR.RE.3	Trusted Advisor	Provide support and guidance to MSE

With our shift to the use side of the IUA, we can see the relevant KPIs also largely relate to how users experience and use GEIGER, and how well we have succeeded in raising awareness of GEIGER among both MSEs and (potential) CSDs.

Table 21: KPIs related to translation inference actions.

Number	Description	Work Package
KPI 1.2	Understanding the GEIGER Indicator by CEOs of MSEs ≥ 4.0 on the Mean Opinion Score scale ranging from 1 (bad) to 5 (excellent)	WP4
KPI 5.3	Satisfaction by MSEs using the GEIGER Framework ≥ 4.0 on “Mean Opinion Score” scale ranging from 1 (bad) to 5 (excellent)	WP4
Impact KPI I2.1.1.1	>500’000 MSEs will be aware of the GEIGER Indicator as a dynamic risk monitoring instrument	WP5
Impact KPI I2.1.1.2	>50’000 MSEs will have tried the personalised GEIGER Indicator for their own specific MSE&ME by registering on GEIGER Solution	WP5
Impact KPI I2.1.1.3	>50 MSEs will have benefitted from the Security Defender education in the Swiss pilot performed with apprentices by the school BBB	WP4

Impact KPI I2.1.1.4	>50 start-up MSEs will have benefitted from the Security Defender education in the Romanian pilot performed by the incubator/accelerator CLUJ IT CLUSTER	WP4
Impact KPI I2.1.1.5	>370 MSEs will have benefitted from advice by an accountant with Security Defender education in the Dutch pilot performed by the education provider ULEI	WP4
Impact KPI I2.1.1.9	GEIGER capacity-building assessed in surveys with >1'000 responses	WP4
Impact KPI I2.1.1.10	GEIGER capacity-building refined in >10 events targeting MSEs.	WP4
Impact KPI I2.1.2.1	perceived level transparency of risks ≥ 4.0 on MOS scale	WP4
Impact KPI I2.1.2.2	perceived level of decision support for risk reduction ≥ 4.0 on MOS scale	WP4
Impact KPI I2.1.2.3	perceived level of risk explanation ≥ 4.0 on MOS scale	WP4
Impact KPI I2.1.3.1	at least 80% of basic recommendations for human error prevention are adopted by the pilot MSEs	WP4
Impact KPI I2.1.3.2	Shield tools are available and in use by the pilot MSEs for protecting against at least 80% of attacks recommended for protection by the participating CERTs/CSIRTs.	WP4
Impact KPI I2.1.4.1	≥1'000'000 impressions of the GEIGER Indicator as measured by number of impressions of media channels.	WP5
Impact KPI I2.1.4.2	≥100'000 small enterprises have a GEIGER account, allowing them to predict their risk with the personalised GEIGER Indicator and benefit from the GEIGER toolbox.	WP5
Impact KPI I2.1.5.1	≥1'000 industry-diverse MSEs that know the GEIGER Indicator.	WP4

In the translation inference we rely on the user to understand GEIGER and to be able to translate the outcome to their own situation, possibly with the help of a CSD. Unsurprisingly, the rebuttals for this inference focus on a lack of understanding by the MSE owner, an unwillingness to participate by employees, and a lack of confidence and proactive nature at CSDs and trusted advisors. Many of these rebuttals will need to be addressed with the help of action research, where we can identify the potential pain points in user interactions. This will largely serve as feedback to WP2, to then improve the usability of the application.

Table 22: Rebuttals to translation inference actions.

Label	Action	KPI	Rebuttal Description	Research Method(s)
R.TR.1	A.TR.VI.1	I2.1.2.1, I2.1.2.2, I2.1.2.3, I2.1.3.1	The user is not able to intuitively find recommendations in the GEIGER application.	Action Research, Case Study
R.TR.2	A.TR.VI.2	1.1, I2.1.4.1, I2.1.4.2, I2.1.5.1	The owner is unwilling to perform a GEIGER indicator scan out of privacy concerns.	Action Research, Subject-Based Experiment
R.TR.3	A.TR.VI.3	5.3, I2.1.1.2	Employees and MSE owners turn out to want to interact with GEIGER in different ways that is not accommodated by the application.	Action Research, Subject-Based Experiment
R.TR.4	A.TR.VI.4	5.2, 5.3, I2.1.5.1	GEIGER provides insufficient support for an owner to configure GEIGER for the MSE.	Action Research, Case Study
R.TR.5	A.TR.VI.5	5.2, 5.3	GEIGER provides insufficient support for an employee to configure GEIGER for themselves and their device.	Action Research, Case Study

R.TR.6	A.TR.VI.6		MSE owners do not include all employees in the GEIGER process.	Action Research, Subject-Based Experiment
R.TR.7	A.TR.VI.7	I2.1.4.1	Employees are not willing to take part in GEIGER based on owner's advice.	Action Research, Case Study
R.TR.8	A.TR.VI.8		There is no GEIGER support desk.	Illustrative Scenario, Logical Argument
R.TR.9	A.TR.VI.9		Trusted advisors do not have sufficient GEIGER knowledge to help MSEs with onboarding.	Case Study, Subject-Based Experiment
R.TR.10	A.TR.VI.10	5.2, I2.1.4.2	MSE associations and the GEIGER curator are unable to reach enough MSEs.	Survey
R.TR.11	A.TR.VI.11	I2.1.1.1	The GEIGER consortium inadequately raises awareness of the GEIGER solution.	Survey
R.TR.12	A.TR.RE.1		GEIGER users are reluctant to ask for help from CSDs since they do not know these people.	Action Research, Subject-Based Experiment, Survey
R.TR.13	A.TR.RE.2	I2.1.1.3, I2.1.1.4, I2.1.1.5	CSDs do not feel confident enough in their knowledge of GEIGER to help and train MSEs.	Subject-Based Experiment, Survey
R.TR.14	A.TR.RE.3	I2.1.1.3, I2.1.1.4, I2.1.1.5	Trusted advisors to MSEs do not believe in the ability of the GEIGER solution to help MSEs.	Action Research, Survey

3.4.2.2 Action Inference

The first user action connected to the action inference is, aptly, 'Determine most appropriate action' (DE). We rely on the MSE to be able to select the recommendation that is most suitable for them with the help of a CSD. The second user action is 'Report incident or enact recommendation' (RE). Here, the user must actually take action and either report an incident they are experiencing or enact a recommendation.

Table 23: Actions coupled to the action inference.

Label	Role	Action
A.AC.DE.1	Owner	Determine most appropriate action
A.AC.DE.2	(C)SD	Guide MSE in action evaluation
A.AC.RE.1	Owner	Report incident or enact recommendation

Although the number of actions related to the action inference is small, many KPIs can be related to this inference, since the actions relate to various elements of the GEIGER solution.

Table 24: KPIs related to action inference actions.

Number	Description	Work Package
KPI 1.2	Understanding the GEIGER Indicator by CEOs of MSEs ≥ 4.0 on the Mean Opinion Score scale ranging from 1 (bad) to 5 (excellent)	WP4
KPI 1.3	Predicted attack intensity (e.g., affected devices) for a specific attack matches +/- 20% in more than 60% of all observed time spans retrospectively	WP4
KPI 3.2	4 open APIs allowing connectivity with MSEs, MSE associations, and CERTs/CSIRTs, and third-party tool and framework providers.	WP2

KPI 5.3	Satisfaction by MSEs using the GEIGER Framework ≥ 4.0 on “Mean Opinion Score” scale ranging from 1 (bad) to 5 (excellent)	WP4
Impact KPI I2.1.1.9	GEIGER capacity-building assessed in surveys with >1’000 responses	WP4
Impact KPI I2.1.1.10	GEIGER capacity-building refined in >10 events targeting MSEs.	WP4
Impact KPI I2.1.2.1	perceived level transparency of risks ≥ 4.0 on MOS scale	WP4
Impact KPI I2.1.2.2	perceived level of decision support for risk reduction ≥ 4.0 on MOS scale	WP4
Impact KPI I2.1.2.3	perceived level of risk explanation ≥ 4.0 on MOS scale	WP4
Impact KPI I2.1.2.5	≥3 CERTs/CSIRTs have access to the incident database	WP2
Impact KPI I2.1.2.6	≥3 data protection authorities have access to the incident database	WP2
Impact KPI I2.1.2.7	≥150 Security Defenders have access to the incident database	WP2
Impact KPI I2.1.2.8	1 open API with API access governance policies for querying incidents and submitting information	WP5
Impact KPI I2.1.2.9	4 contributions to standardisation work or MoUs with related initiatives for harmonising external GEIGER Framework interfaces and the security defenders education.	WP5
Impact KPI I2.1.3.1	at least 80% of basic recommendations for human error prevention are adopted by the pilot MSEs	WP4
Impact KPI I2.1.3.2	Shield tools are available and in use by the pilot MSEs for protecting against at least 80% of attacks recommended for protection by the participating CERTs/CSIRTs.	WP4
Impact KPI I2.1.3.3	90% of the incidents experienced by the pilot MSEs are detected and resolved within 30 days	WP4

Currently we have only coupled one rebuttal to each action for this inference. Nevertheless, there are many questions one can ask about the user and CSD actions of this inference. This can be seen partially in the number of KPIs that could relate to each rebuttal described in Table 25. It is likely that several warrants and/or backings will be required to address these rebuttals.

Table 25: Rebuttals to action inference actions.

Label	Action	KPI	Rebuttal Description	Research Method(s)
R.AC.1	A.AC.DE.1	1.2, 5.3, I2.1.1.9, I2.1.1.10, I2.1.3.1	MSEs are not willing to take any action since they feel the costs outweigh the benefits.	Survey
R.AC.2	A.AC.DE.2	I2.1.2.7	Although their knowledge of GEIGER is sufficient, CSDs are not able to adapt to specific situations to provide actionable advice for different types of MSEs.	Action Research, Subject-Based Experiment
R.AC.3	A.AC.RE.1	I2.1.2.9, I2.1.3.1, I2.1.3.3	Users do not trust GEIGER enough to report incidents.	Case Study, Survey

3.4.2.3 Reflection Inference

The final inference is the reflection inference, where the only coupled user action is ‘Receive and process feedback’ (RE). To internalize the feedback, the MSE may request the help of a trusted advisor.

Table 26: Actions coupled to the reflection inference.

Label	Role	Action
A.RE.RE.1	Owner	Receive and process feedback
A.RE.RE.2	Trusted Advisor	Motivate MSE to keep improving

Once more, although few actions are linked to this inference, many KPIs could be of relevance. The opinion of the user regarding the GEIGER application will be highly influential in whether they continue their use of the GEIGER application over a longer period. Only then will the incorporation of feedback start to yield benefits.

Table 27: KPIs related to reflection inference actions.

Number	Description	Work Package
KPI 1.2	Understanding the GEIGER Indicator by CEOs of MSEs ≥ 4.0 on the Mean Opinion Score scale ranging from 1 (bad) to 5 (excellent)	WP4
KPI 1.3	Predicted attack intensity (e.g., affected devices) for a specific attack matches +/- 20% in more than 60% of all observed time spans retrospectively	WP4
KPI 5.3	Satisfaction by MSEs using the GEIGER Framework ≥ 4.0 on "Mean Opinion Score" scale ranging from 1 (bad) to 5 (excellent)	WP4
Impact KPI I2.1.1.9	GEIGER capacity-building assessed in surveys with >1'000 responses	WP4
Impact KPI I2.1.1.10	GEIGER capacity-building refined in >10 events targeting MSEs.	WP4
Impact KPI I2.1.2.1	perceived level transparency of risks ≥ 4.0 on MOS scale	WP4
Impact KPI I2.1.2.2	perceived level of decision support for risk reduction ≥ 4.0 on MOS scale	WP4
Impact KPI I2.1.2.3	perceived level of risk explanation ≥ 4.0 on MOS scale	WP4
Impact KPI I2.1.3.1	at least 80% of basic recommendations for human error prevention are adopted by the pilot MSEs	WP4
Impact KPI I2.1.3.3	90% of the incidents experienced by the pilot MSEs are detected and resolved within 30 days	WP4

As with the previous inference, the rebuttals related to the relevant actions for this inference may require several warrants and/or backings to be properly addressed. Since we are additionally discussing the topic of 'continued use' these experiments may need to be carried out over an extended period.

Table 28: Rebuttals to reflection inference actions.

Label	Action	KPI	Rebuttal Description	Research Method(s)
R.RE.1	A.RE.RE.1	1.2, 5.3, I2.1.1.9, I2.1.1.10, I2.1.2.1, I2.1.2.2, I2.1.2.3	MSEs do not feel the need for continued use of GEIGER after an initial scan and implementation of recommendations.	Action Research, Subject-Based Experiment, Survey
R.RE.2	A.RE.RE.2	1.3, 5.3, I2.1.1.9, I2.1.1.10	The trusted advisor is no longer convinced of the GEIGER solution and does not motivate the MSE user.	Subject-Based Experiment, Survey

3.5 Conclusion

In this section we introduced our theoretical validation framework. After discussing the GEIGER journey and personas (3.1), we covered the Hopster-den Otter (2019) validation framework used in validation for (embedded) formative assessment (3.2). We detailed the principles of argument-based validation (3.3), touching on Toulmin arguments (3.3.1) and the GEIGER argumentation chain (3.3.2). Finally, we discussed our extensive argumentation guidelines, providing related actions, KPIs, rebuttals, and research methods for each inference in the GEIGER IUA chain.

Although this section has brought us quite far from a theoretical standpoint, the practical planning and implementation of validation activities remains to be discussed. In the next section, we will detail how we used the results obtained here to formulate a validation planning for the GEIGER project. Then, in Section 5, we will cover validation activities that have taken place during the first months of WP4.

3.6 References

1. Boletsis, C., Halvorsrud, R., Pickering, J. B., Phillips, S. C., & Surridge, M. (2021, February). Cybersecurity for MSEs: Introducing the Human Element into Socio-technical Cybersecurity Risk Assessment. In *VISIGRAPP (3: IVAPP)* (pp. 266-274).
2. Cook, T. D., Campbell, D. T., & Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
3. Cook, D. A., Brydges, R., Ginsburg, S., & Hatala, R. (2015). A contemporary approach to validity arguments: a practical guide to K ane's framework. *Medical education*, 49(6), 560-575.
4. Cooper, R., & Foster, M. (1971). Sociotechnical systems. *American Psychologist*, 26(5), 467.
5. Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, 52(4), 281.
6. Davis, M. C., Challenger, R., Jayewardene, D. N., & Clegg, C. W. (2014). Advancing socio-technical systems thinking: A call for bravery. *Applied ergonomics*, 45(2), 171-180.
7. Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science education*, 88(6), 915-933.
8. Fenz, S., & Ekelhart, A. (2010). Verification, validation, and evaluation in information security risk management. *IEEE Security & Privacy*, 9(2), 58-65.
9. Halvorsrud, R., Kvale, K., & Følstad, A. (2016a). Improving service quality through customer journey analysis. *Journal of service theory and practice*, 26(6), 840-867.
10. Halvorsrud, R., Haugstveit, I. M., & Pultier, A. (2016b, September). Evaluation of a modelling language for customer journeys. In *2016 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)* (pp. 40-48). IEEE.
11. Halvorsrud, R., Lillegaard, A. L., Røhne, M., & Jensen, A. M. (2019). Managing complex patient journeys in healthcare. In *Service Design and Service Thinking in Healthcare and Hospital Management* (pp. 329-346). Springer, Cham.
12. Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS quarterly*, 75-105.
13. Hopster-den Otter, D., Wools, S., Eggen, T. J., & Veldkamp, B. P. (2019). A general framework for the validation of embedded formative assessment. *Journal of educational measurement*, 56(4), 715-732.
14. Kane, M. T. (1992). An argument-based approach to validity. *Psychological bulletin*, 112(3), 527.
15. Kane, M. (2013). The argument-based approach to validation. *School Psychology Review*, 42(4), 448-457.
16. Mingers, J., & Standing, C. (2020). A framework for validating information systems research based on a pluralist account of truth and correctness. *Journal of the Association for Information Systems*, 21(1), 6.
17. Paja, E., Dalpiaz, F., & Giorgini, P. (2015). Modelling and reasoning about security requirements in socio-technical systems. *Data & Knowledge Engineering*, 98, 123-143.

18. Peffers, K., Rothenberger, M., Tuunanen, T., & Vaezi, R. (2012, May). Design science research evaluation. In *International Conference on Design Science Research in Information Systems* (pp. 398-410). Springer, Berlin, Heidelberg.
19. Pries-Heje, J., Baskerville, R., & Venable, J. R. (2008). Strategies for design science research evaluation.
20. Recker, J., Zur Muehlen, M., Siau, K., Erickson, J., & Indulska, M. (2009). Measuring method complexity: UML versus BPMN. In *Proceedings of the Fifteenth Americas Conference on Information Systems* (pp. 1-9). Association for Information Systems.
21. Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action design research. *MIS quarterly*, 37-56.
22. Simon, S. (2008). Using Toulmin's argument pattern in the evaluation of argumentation in school science. *International Journal of Research & Method in Education*, 31(3), 277-289.
23. Shojaifar, A., Fricker, S. A., & Gwerder, M. (2020, September). Automating the Communication of Cybersecurity Knowledge: Multi-case Study. In *IFIP World Conference on Information Security Education* (pp. 110-124). Springer, Cham.
24. Stockdale, R., & Standing, C. (2006). An interpretive approach to evaluating information systems: A content, context, process framework. *European journal of operational research*, 173(3), 1090-1102.
25. Straub, D. W. (1989). Validating instruments in MIS research. *MIS quarterly*, 147-169.
26. Straub, D., Boudreau, M. C., & Gefen, D. (2004). Validation guidelines for IS positivist research. *Communications of the Association for Information systems*, 13(1), 24.
27. Toulmin, S. E. (2003). *The uses of argument*. Cambridge university press.
28. Venable, J., Pries-Heje, J., & Baskerville, R. (2012, May). A comprehensive framework for evaluation in design science research. In *International conference on design science research in information systems* (pp. 423-438). Springer, Berlin, Heidelberg.
29. Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: a framework for evaluation in design science research. *European journal of information systems*, 25(1), 77-89.
30. Wieringa, R. J. (2014). *Design science methodology for information systems and software engineering*. Springer.
31. Wools, S., Sanders, P., & Eggen, T. (2010). Evaluation of validity and validation by means of the argument-based approach. *Evaluation of Validity and Validation by Means of the Argument-based Approach*, 1000-1020.

4 Validation Planning

The rebuttals presented in Section 3 are by no means a final list. Over time, as we conduct validation activities and ascertain different issues in different use case countries, the list of rebuttals will grow. Nevertheless, the rebuttals we currently have serve as an excellent baseline from which to construct our initial validation planning.

We presented a total of 68 rebuttals in Section 3. The simplest planning approach would assign a research activity to each rebuttal to investigate whether its claims are correct and whether a warrant or backing can be unearthed to counter the rebuttal. However, this would imply planning 68 separate research activities. Hence, we sought to combine similar research activities where possible and grouped sets of research activities related to the same target user groups in validation phases.

Before discussing our validation phases in Section 4.2, we will first discuss some overall project timelines which helped to form our WP4 timelines in Section 4.1. In Section 4.2, we will first cover the activities conducted during our preparatory phase. Then, we will present our planning for the three main phases of validation. Lastly, in Section 4.3, we will discuss how we expect to remain flexible in our planning in the future, without losing structure in our validation approach.

4.1 Project Timelines

Our plans for validation must be aligned with the progress made by the GEIGER project. It does not make sense to plan an experiment involving a fully integrated tool at a time before the technical work package WP2 intends to have this integrated tool available. Similarly, it does not make sense to plan validation activities involving certified security defenders (CSDs) at times earlier than the education work package WP4 plans to have the first CSDs educated and certified.

Table 29 indicates for all work packages except WP4 at which points in the project which deliverables are due. This helps us to get a sense of when it is possible to plan certain validation activities. We have indicated in yellow those deliverables which directly influence our validation planning in the period M13-M30.

Table 29: GEIGER deliverables and their timings. Deliverables directly influencing validation planning are indicated in yellow.

Work Package	M01-M06	M12	M18	M24	M30
WP1	Requirements	Architecture			
WP2		Adapted Components	Integrated Prototype	Framework MVP	Framework Release
WP3	Training Plan		Intermediate Training Report		Final Training Report
WP5	Impact Plan		Intermediate Impact Report		Final Impact Report
WP6	Data Management Plan	Year 1 Report		Year 2 Report	Final Report
WP7	Research Participant Selection, Personal Data Protection				

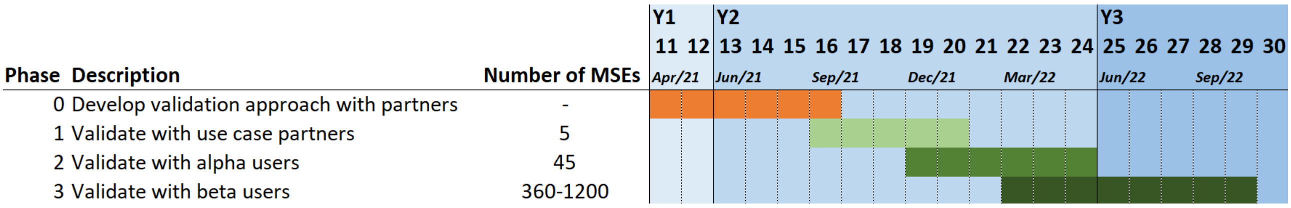
We can see from Table 29 that WP2 impacts WP4 most in this dimension. The work for the 'Adapted Components' deliverable that was due at M12 was decisive in determining what was available for validation from a technical sense in the first months of WP4. The period between the 'Adapted Components' deliverable and the 'Integrated Prototype' deliverable marks a time in validation where we do not yet have an integrated tool with which to perform experiments with users. For us, this meant our focus in user testing in this phase would be on interactions with the GEIGER use case partners. Already involving alpha partners at this early stage is premature since we would not have a working application to let them interact with.

Month M18 marks a turning point in the GEIGER project. The 'Integrated Prototype' deliverable is due in this month. Additionally, the 'Intermediate Training Report' from WP3 and the 'Intermediate Impact Report' from WP5 are due. This means that from month M19 onwards, we can expect to work with an integrated GEIGER prototype. The progress on the education front in terms of CSD training and the progress on the dissemination front in terms of attracting potential users mean that we can increase the scope of our validation efforts. From month M19 we can start to include alpha users in our validation planning and can also begin experiments involving CSDs.

The 'Framework MVP' deliverable due in month M24 marks a final turning point in validation. From M24 onwards we can expect to have a functioning minimum viable product (MVP). Along with the progress made on the education and dissemination fronts, this implies we can start to include beta users to validate the GEIGER solution. The beta user group will be much larger than the alpha user group, and we will therefore have fewer possibilities for interaction. This means it is vital to have a vibrant CSD community to guide users at times that the GEIGER consortium cannot. Additionally, since we require a large group of beta users and these users need to be recruited via indirect networks and the GEIGER multipliers, we need to start the recruitment phase well before month M24. In Section 4.2, we will discuss how our validation phases align with the observations we have formulated regarding the GEIGER project timelines in this section.

4.2 Validation Phases

Figure 20 depicts the validation phases resulting from our analysis in Section 4.1. The preparations for validation started before the official commencement of WP4 in month M13. Given our theoretical approach to validation, we used this extra time to form our validation concept and communicate it to the relevant partners in the consortium. From month M13 to M16, we refined our theoretical framework and initiated our validation planning efforts. This culminated in reaching an agreement on our overall validation planning in M16.



			Y1		Y2												Y3					
			11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Phase	**Description**	**Number of MSEs**	Apr/21		Jun/21			Sep/21			Dec/21			Mar/22			Jun/22			Sep/22		
0	Develop validation approach with partners	-	🟧	🟧	🟧	🟧	🟧	🟧														
1	Validate with use case partners	5						🟩	🟩	🟩	🟩	🟩										
2	Validate with alpha users	45									🟩	🟩	🟩	🟩	🟩	🟩						
3	Validate with beta users	360-1200												■	■	■	■	■	■	■	■	

Figure 20: GEIGER validation phases.

Our modular, stepwise approach to validation aims to ensure that we are flexible enough to adapt to unexpected circumstances, while still ensuring to perform all necessary validation activities. Given that the integrated GEIGER prototype is available from month M19, we focus our initial validation efforts (Phase 1) on interactions with the GEIGER use case partners and related experiments.

From month M19, with the availability of an integrated prototype and the necessary education and dissemination support structure in place, we can begin validation Phase 2: validating with alpha users. This constitutes a group of roughly 45 MSEs, corresponding to 15 MSEs per GEIGER use case country. These MSEs are intended to represent the various personas we identified as relevant during the preparatory phase.

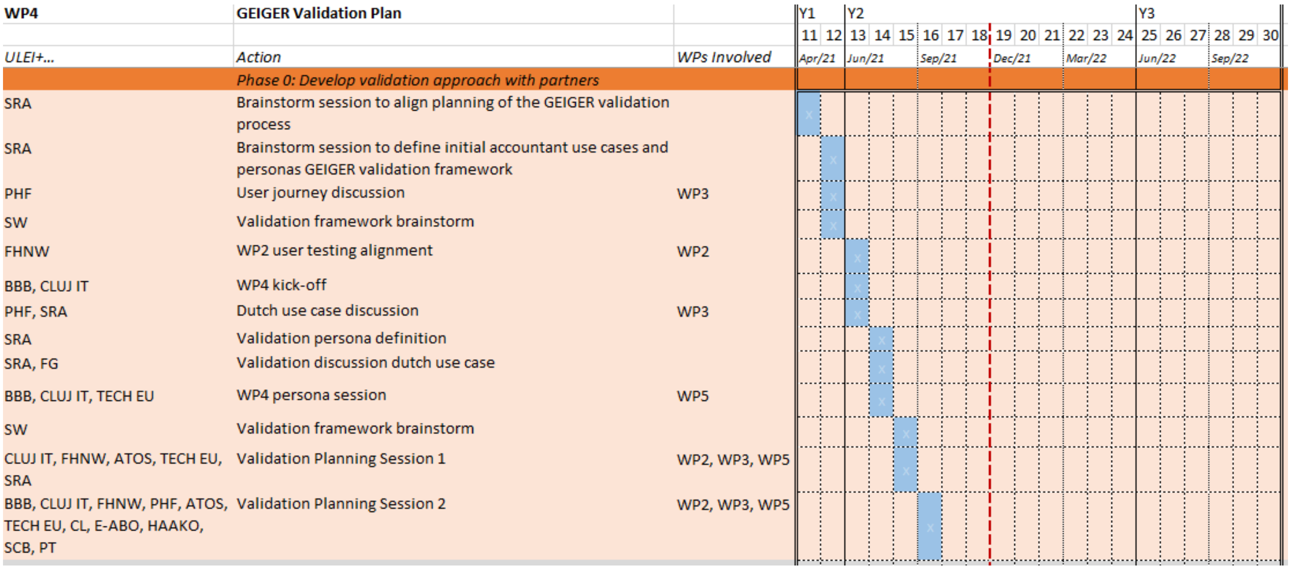
The final phase of validation (Phase 3) commences in month M22, with concerted dissemination and recruitment efforts together with WP5 and the GEIGER multipliers. The goal is to eventually reach a group of 360-1200 beta users to test the GEIGER application in the months M24-M29.

In the following sections we will provide more details on each validation phase and exactly which activities are currently planned to take place.

4.2.1 Preparatory Phase

In preparation for the validation work package, we conducted several brainstorm sessions and had various discussions to form our theoretical ideas and turn them into practical implementations. Figure 21 shows the activities we conducted, the GEIGER partners that were involved, the GEIGER work packages that were involved, and the timelines of the activities. We can already observe that WP4 work involves interaction with many other GEIGER work packages.

In month M11 (April of 2021) we performed the first estimations of validation scenarios and timelines, with SRA as the launching partner in The Netherlands. We further refined our ideas in month M12. We aligned on the envisioned use of GEIGER by accountants and their client MSEs. We additionally participated in WP3 discussions on user journeys and aligned on GEIGER support infrastructure ideas.



WP4	GEIGER Validation Plan		Y1		Y2												Y3					
			11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
ULEI+...	*Action*	*WPs Involved*	*Apr/21*		*Jun/21*			*Sep/21*			*Dec/21*			*Mar/22*			*Jun/22*			*Sep/22*		
	Phase 0: Develop validation approach with partners																					
SRA	Brainstorm session to align planning of the GEIGER validation process		x																			
SRA	Brainstorm session to define initial accountant use cases and personas GEIGER validation framework			x																		
PHF	User journey discussion	WP3		x																		
SW	Validation framework brainstorm			x																		
FHNW	WP2 user testing alignment	WP2			x																	
BBB, CLUJ IT	WP4 kick-off				x																	
PHF, SRA	Dutch use case discussion	WP3			x																	
SRA	Validation persona definition					x																
SRA, FG	Validation discussion dutch use case					x																
BBB, CLUJ IT, TECH EU	WP4 persona session	WP5				x																
SW	Validation framework brainstorm						x															
CLUJ IT, FHNW, ATOS, TECH EU, SRA	Validation Planning Session 1	WP2, WP3, WP5					x															
BBB, CLUJ IT, FHNW, PHF, ATOS, TECH EU, CL, E-ABO, HAAKO, SCB, PT	Validation Planning Session 2	WP2, WP3, WP5						x														

Figure 21: Preparatory validation phase. FG indicates Frank Grimberg of FHNW. SW indicates Saskia Wools of Cito.

Month M12 also marked our first interaction with Saskia Wools, who is the Head of the Research, Knowledge, and Innovation department at Cito, the Dutch institute for the development of exams and tests. Saskia Wools has extensive experience in validation, specifically in the educational measurement domain. She contributed to the Hopster-den Otter et al. (2019) work as a co-author; the work we based our theoretical validation framework on. She has also written many other scientific contributions in this area. She has supported us in developing our theoretical validation framework, helping to establish a firm basis for all WP4 work.

In month M13, official work for WP4 started. We had an alignment meeting on progress made so far in WP2 with user testing for the UI design process in GEIGER. We additionally had a WP4 kick-off meeting with the WP4 task leaders, to discuss our theoretical validation framework and the need for persona definitions. We also continued our collaboration with WP3, by meeting with SRA and PHF to discuss the details of the Dutch use case from both the WP3 and WP4 perspective.

Month M14 involved various meetings to take the first concrete steps in persona definition. This included a session on location at SRA, with Frank Grimberg of FHNW, which resulted in the first complete definition of personas for the Dutch use case. We also involved WP5 more actively in this phase, to ensure that dissemination activities are aligned with the personas we identified as relevant for WP4.

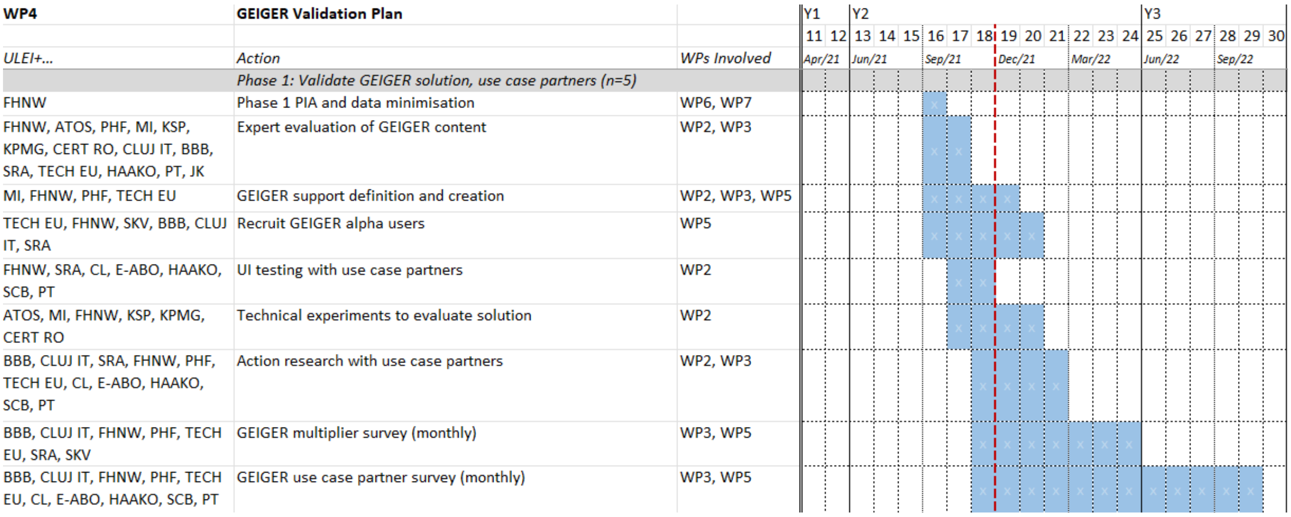
In months M15 and M16 we had a session with Saskia Wools to iron out the details of our theoretical validation framework and used the results as input for validation planning sessions with all WP4 partners. We actively involved the use case partners in these discussions since they would be involved in the first phase of validation. This marked the end of our validation preparation phase, meaning the first validation activities of Phase 1 could begin.

4.2.2 Phase 1: Validate with Use Case Partners

The first phase of validation involves various experiments with the GEIGER use case partners. Additionally, we must ensure the necessary measures are in place to ensure correct data processing in this phase, and to ensure that we are sufficiently prepared for further validation phases.

With help from FHNW and the supporting documentation from WP6 and WP7, we conducted a data protection impact assessment (DPIA) for the Phase 1 activities where this was considered necessary. For more details on these assessments, please consult Section 6.

In months M16 and M17, several activities took place to evaluate the GEIGER content with an expert panel. Experts from within the GEIGER consortium performed an evaluation of the GEIGER content to assess how well it addresses the security, privacy, and GDPR focus areas of the GEIGER project. The aim of this expert panel was to elicit areas within the project where our validation efforts should be focused. Essentially, these are the areas where rebuttals are the strongest, and where argumentation and assumptions are the weakest. The detailed results of the expert panel can be found in Section 5.1.1.



WP4	GEIGER Validation Plan		Y1		Y2												Y3					
			11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
ULEI+...	*Action*	*WPs Involved*	*Apr/21*		*Jun/21*			*Sep/21*			*Dec/21*			*Mar/22*			*Jun/22*			*Sep/22*		
	Phase 1: Validate GEIGER solution, use case partners (n=5)																					
FHNW	Phase 1 PIA and data minimisation	WP6, WP7						x														
FHNW, ATOS, PHF, MI, KSP, KPMG, CERT RO, CLUJ IT, BBB, SRA, TECH EU, HAAKO, PT, JK	Expert evaluation of GEIGER content	WP2, WP3						x	x													
MI, FHNW, PHF, TECH EU	GEIGER support definition and creation	WP2, WP3, WP5						x	x	x	x											
TECH EU, FHNW, SKV, BBB, CLUJ IT, SRA	Recruit GEIGER alpha users	WP5						x	x	x	x	x										
FHNW, SRA, CL, E-ABO, HAAKO, SCB, PT	UI testing with use case partners	WP2							x	x												
ATOS, MI, FHNW, KSP, KPMG, CERT RO	Technical experiments to evaluate solution	WP2							x	x	x	x										
BBB, CLUJ IT, SRA, FHNW, PHF, TECH EU, CL, E-ABO, HAAKO, SCB, PT	Action research with use case partners	WP2, WP3								x	x	x	x									
BBB, CLUJ IT, FHNW, PHF, TECH EU, SRA, SKV	GEIGER multiplier survey (monthly)	WP3, WP5								x	x	x	x	x	x	x						
BBB, CLUJ IT, FHNW, PHF, TECH EU, CL, E-ABO, HAAKO, SCB, PT	GEIGER use case partner survey (monthly)	WP3, WP5								x	x	x	x	x	x	x	x	x	x	x	x	

Figure 22: Phase 1 of validation.

Months M16 and M17 also saw the first efforts to define and create the GEIGER support function, in collaboration with WP2, WP3, and WP5. These efforts are intended to continue in months M18 and M19. We also saw the first efforts from all use case country leads to recruit GEIGER alpha users for Phase 2 of validation.

In these months the first technical experiments in task T2.5 on Optimisation and Hardening took place. The technical experiments should demonstrate the proper functioning of the integrated GEIGER prototype, towards the 'Integrated Prototype' deliverable due for WP2. Although these results will largely be incorporated in WP2 work, they have relevance to our activities in WP4, as we saw in certain rebuttals presented in Section 3.4.

From month M17 onwards, we have been involved in experimentation involving user interactions. Action research with use case partners first involved user testing with GEIGER click-dummies. This work is described in Section 5.1.2. To create clarity in our planning, we separated these first UI tests from remaining action research, where we intend to interact more intensively with users. The two activities were considered concurrently in our DPIA process described in Section 6.

In the future, testing will involve increasingly advanced GEIGER prototypes, as we work towards using the integrated prototype in testing. Our action research will be used to uncover how the use case partners interact with the GEIGER solution. This includes how they proceed in motivating employees to use the application, determining what role trusted advisors play in the process, finding out how independent the user is in configuring and using GEIGER, eliciting unexpected interactions with the tool, and observing privacy issues the MSEs experience.

Finally, we have also initiated surveys with the GEIGER multipliers and the use case partners. The GEIGER multipliers SRA, CLUJ IT, and SKV (in combination with BBB) receive a monthly survey to track their progress in motivating MSEs to use GEIGER and reaching new candidate MSEs for later validation phases. This process involves collaboration with WP5. The use case partners receive a different and more extensive monthly survey. We measure their progress from the moment they start using the GEIGER solution. Questions relate to understanding, technical issues, motivational issues, relevance of content, and much more. For more details on the surveys, see Section 5.1.1.

Altogether, the results of Phase 1 of validation will help to address many of the rebuttals mentioned in Section 3.4. Nevertheless, we are only experimenting with use case partners, who have had an active role in developing GEIGER. The next step is to validate GEIGER with a broader and more critical audience.

4.2.3 Phase 2: Validate with Alpha Users

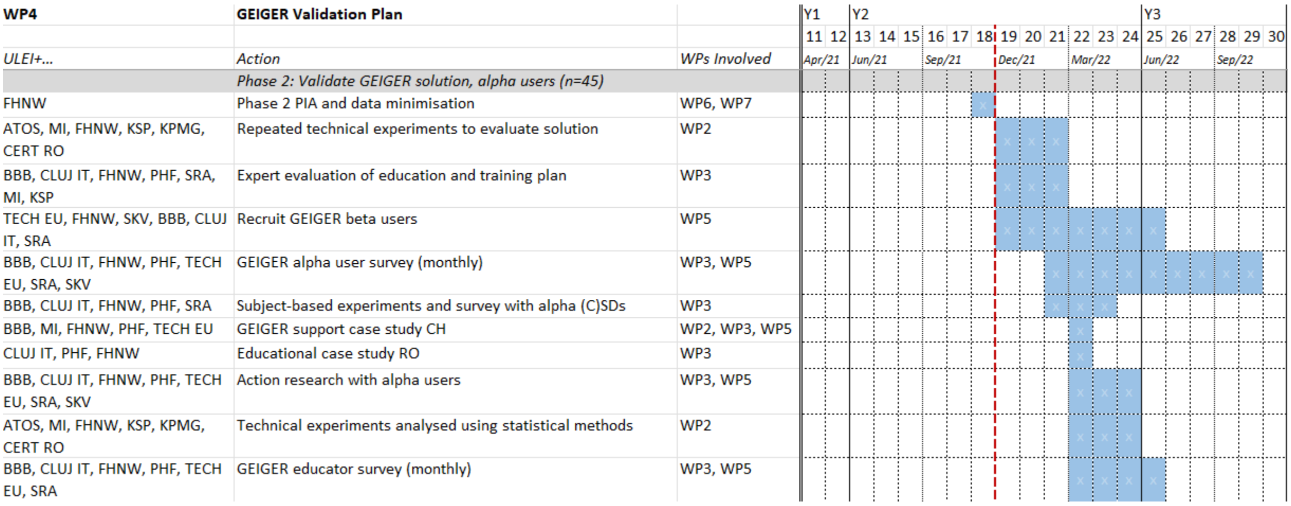
Validation Phase 2 is due to commence in month M19. This implies that the DPIAs for this validation phase should take place before M19. Given the ‘Integrated Prototype’ deliverable deadline at the end of M18, we intend to start various validation activities in M19.

First, repeated technical experiments performed by T2.5 should help to evaluate the integrated solution. Earlier experiments should have shown proper functioning of the GEIGER solution. We now want to study the functioning over time. Do integrated tools update their data often enough? Does the toolbox store data for a sufficiently long period of time? Is enough data built up for an accurate assessment by the GEIGER indicator? How is the SME IT infrastructure included in the picture? Does the cloud push updates to the toolbox successfully? The answers to all these questions, and more, will serve as invaluable input for both WP2 and WP4.

Month M19 also sees the start of another expert evaluation, this time of the education and training plan which will be described in detail in the ‘Intermediate Training Report’ deliverable of WP3. The GEIGER education and training plan should be evaluated in interactive sessions between the creators and the users (educators) of the plan. The educators will be able to make clear any worries (such as lack of translation into native language, disagreements on the requirements for certification, etc.). The creators will be able to argue how they will address any issues. The findings can serve as input to determine the types of questions to ask in future education-related research.

Starting in month M19, a concerted effort by the whole consortium is required to recruit enough MSEs to participate as beta users in Phase 3 of GEIGER validation. Somewhere between 360 and 1,200 MSEs are needed, spread across the three use case countries Switzerland, The Netherlands, and Romania. The official GEIGER launch event is due to take place in the period M23-M25 and will be of great help in this regard. We additionally intend to have at least one GEIGER beta recruitment event in each use case country in the period M22-M25.

From month M21, we will conduct a monthly survey of the GEIGER alpha users. We will ask alpha users questions regarding their experience with the GEIGER solution. It will be less extensive than the use case partner survey. It will focus on issues such as which features the alpha users like and do not like, why they would buy the GEIGER product, who they would recommend GEIGER to, etc. Besides this, we would include questions also contained in the use case partner survey on the current situation of the MSE. Did they recently experience an incident? How do they see their current cybersecurity posture? Etc.



WP4	GEIGER Validation Plan		Y1		Y2												Y3					
			11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
ULEI+...	*Action*	*WPs Involved*	*Apr/21*		*Jun/21*			*Sep/21*			*Dec/21*			*Mar/22*			*Jun/22*			*Sep/22*		
	Phase 2: Validate GEIGER solution, alpha users (n=45)																					
FHNW	Phase 2 PIA and data minimisation	WP6, WP7								x												
ATOS, MI, FHNW, KSP, KPMG, CERT RO	Repeated technical experiments to evaluate solution	WP2									x	x	x									
BBB, CLUJ IT, FHNW, PHF, SRA, MI, KSP	Expert evaluation of education and training plan	WP3									x	x	x									
TECH EU, FHNW, SKV, BBB, CLUJ IT, SRA	Recruit GEIGER beta users	WP5									x	x	x	x	x	x	x					
BBB, CLUJ IT, FHNW, PHF, TECH EU, SRA, SKV	GEIGER alpha user survey (monthly)	WP3, WP5											x	x	x	x	x	x	x	x	x	
BBB, CLUJ IT, FHNW, PHF, SRA	Subject-based experiments and survey with alpha (C)SDs	WP3											x	x	x							
BBB, MI, FHNW, PHF, TECH EU	GEIGER support case study CH	WP2, WP3, WP5												x								
CLUJ IT, PHF, FHNW	Educational case study RO	WP3												x								
BBB, CLUJ IT, FHNW, PHF, TECH EU, SRA, SKV	Action research with alpha users	WP3, WP5												x	x	x						
ATOS, MI, FHNW, KSP, KPMG, CERT RO	Technical experiments analysed using statistical methods	WP2												x	x	x						
BBB, CLUJ IT, FHNW, PHF, TECH EU, SRA	GEIGER educator survey (monthly)	WP3, WP5												x	x	x	x					

Figure 23: Phase 2 of validation.

The first people undergoing alpha education to become a CSD, will be surveyed in months M21-M23 and will participate in subject-based experiments. This will allow us to answer important questions such as: Are potential CSDs willing to become involved in GEIGER education? Are the education and training sessions accessible to all CSDs? Do CSDs feel the necessity to take trainings? Or do they feel they already know enough? Do potential CSDs want to become certified? Do potential CSDs believe in the ability of GEIGER to improve cybersecurity awareness and resilience? All these questions will likely have different answers for different personas.

In this phase of validation, we should also perform a case study of an MSE user contacting GEIGER support, potentially via a CSD. We intend to conduct this case study in Switzerland. The goal is to demonstrate the capability of GEIGER support to deal with such requests. At this stage, we should also have a clear definition of the support task and logistical issues such as availability of support. In the same month M22, we intend to conduct an educational case study in Romania in collaboration with WP3. The case study will walk through all steps of the educational process. From initial training of CSDs, to certification, to CSD matchmaking, to CSDs helping the MSE. This should help to answer the following questions: Are MSE associations able to perform the matchmaking? Are there barriers we did not envision beforehand? Do educators have enough knowledge?

In month M22 we initiate various other validation activities. We will conduct action research with alpha users. We will observe how users interact with the GEIGER solution and actively discuss with them what problems they specifically run into. This will allow us to discover any unexpected user journey deviations for specific GEIGER persona, that we will not have been able to discover at an earlier stage. Here we should also have some CSD structure, so that we can observe if people feel comfortable asking for help. An important question is whether users trust GEIGER, for example to report incidents.

Further technical experiments should be performed in T2.5, potentially analysed using statistical methods. These experiments should be carried out in conjunction with the action research on alpha users. This should help to answer some of the remaining questions regarding performance of the GEIGER solution. Do the GEIGER APIs function properly, also for incident recording? Does the toolbox adapt based on user input and indications of implemented recommendations? Does the cloud connection work and do users enable it? Does updating work? Does pairing and sharing work? Is the GEIGER indicator algorithm updated based on user input?

Finally, we plan to conduct a survey among the GEIGER educators. The main GEIGER educators SRA, CLUJ IT, and BBB will receive a survey to track their experiences with using the GEIGER education and training plan, experiences with trainer trainings, as well as with recruiting and educating CSDs. This will likely involve intense collaboration with PHF (WP3) and TECH EU (WP5). It should also help to determine whether educators are interpreting the plan correctly.

4.2.4 Phase 3: Validate with Beta Users

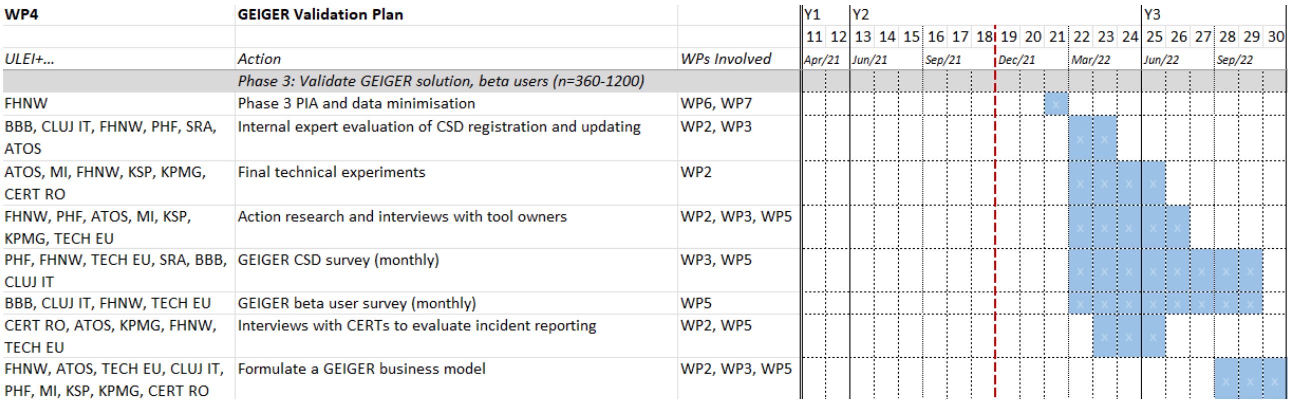
The final phase of validation involves a much larger group of users, with whom we interact much less intensively. Validation Phase 3 intends to address the remaining rebuttals through statistically robust analyses of data collected from large groups of users. Additionally, we intend to address final worries regarding retained relevance of the GEIGER solution over longer time periods. The final work turns the attention to the feasibility of exploitation of the GEIGER solution, where WP4 results provide an important input.

As with the other validation phases, Phase 3 begins with a DPIA to uncover any necessary measures to be taken to ensure sufficient data protection. Then, in month M22, the first experiments will start.

We will perform an expert evaluation of CSD registration and updating. Experts internal to the GEIGER consortium should evaluate whether the CSD registration and retraining procedure functions as planned. Input from the final technical experiments will allow us to determine whether it functions in a technical sense. Is there a program in place to update CSD knowledge after some time? Can we confirm that we know of all registered CSDs (nobody missing, nobody incorrectly registered)?

The final technical experiments should give clarity on potential long-term issues that could not be identified at earlier stages. Does the CSD directory function properly? Does pairing and sharing function at a larger scale? Are there any problems that arise as new tools are included in the solution? Are all previously identified bugs fixed?

Action research and interviews with tool owners is another element commencing in month M22. Active interaction with (new) tool owners, for example via connectathons, should provide insight into what is, and is not, working regarding integrated tools. Do educational tools find the guidelines provided by the educational environment sufficient? What do tool owners feel are reasonable financial plans for MSE users to use their tool? Does this match with the opinion of MSEs? Do tool owners find the API usable? Are tool owners committed to updating their tools regularly?



WP4	GEIGER Validation Plan		Y1		Y2												Y3					
			11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
ULEI+...	*Action*	*WPs Involved*	*Apr/21*		*Jun/21*			*Sep/21*			*Dec/21*			*Mar/22*			*Jun/22*			*Sep/22*		
	Phase 3: Validate GEIGER solution, beta users (n=360-1200)																					
FHNW	Phase 3 PIA and data minimisation	WP6, WP7											x									
BBB, CLUJ IT, FHNW, PHF, SRA, ATOS	Internal expert evaluation of CSD registration and updating	WP2, WP3												x	x							
ATOS, MI, FHNW, KSP, KPMG, CERT RO	Final technical experiments	WP2												x	x	x	x					
FHNW, PHF, ATOS, MI, KSP, KPMG, TECH EU	Action research and interviews with tool owners	WP2, WP3, WP5												x	x	x	x	x				
PHF, FHNW, TECH EU, SRA, BBB, CLUJ IT	GEIGER CSD survey (monthly)	WP3, WP5												x	x	x	x	x	x	x	x	
BBB, CLUJ IT, FHNW, TECH EU	GEIGER beta user survey (monthly)	WP5												x	x	x	x	x	x	x	x	
CERT RO, ATOS, KPMG, FHNW, TECH EU	Interviews with CERTs to evaluate incident reporting	WP2, WP5													x	x	x					
FHNW, ATOS, TECH EU, CLUJ IT, PHF, MI, KSP, KPMG, CERT RO	Formulate a GEIGER business model	WP2, WP3, WP5																		x	x	x

Figure 24: Phase 3 of validation.

The two final surveys will be conducted among CSDs and the beta user group. All CSDs should fill in a monthly survey regarding their progress and their GEIGER interactions. CSDs should indicate whether they are still actively involved in GEIGER (no participation in the survey is also assumed as inactive). If they are no longer active, why not? Are CSDs still convinced of the GEIGER solution? Why or why not? Are the CSDs able to adapt to the different situations at different MSEs they may help? Were the supporting certification authority, educators, and GEIGER support available when needed? Ideally, trusted advisors would also fill in these questions, even if they eventually chose not to become certified.

All GEIGER beta users should agree beforehand to filling in a short survey each month for 6 months, so that we can track their progress. Given the fact that not all MSEs will join at the same time and people will take holidays, we have chosen to use 8 months to collect the 6 months' worth of data. Users answer questions such as: Do they feel the need/motivation for continued use? How satisfied are they with the solution? Is the data provided suitable to their situation (e.g., CERT incident reports)? Do users trust GEIGER enough to report incidents? Can they take the actions they want to take, or are there too many financial barriers? Are CSDs available in times of need? As before, we also ask if they experienced any incidents and whether GEIGER helped them to deal with these incidents.

Around month M24 we will additionally conduct interviews with CERTs to evaluate the incident reporting process. This should uncover any final issues regarding incident reporting in both directions. Are the APIs functioning as they should? Do connected CERTs feel the benefits of connecting to GEIGER outweigh the costs? Whether yes or no, why? Would they recommend connecting to other CERTs? Which ones? What can still be improved?

The final step is to formulate a GEIGER business model. The responsibility of this activity rests with WP5, but the results of WP4 activities will inevitably be used in the construction of the model. Using inputs from all the surveys and experiments carried out, the GEIGER consortium should formulate a GEIGER business model. The model must include estimations for the capacity (in FTE) required to keep all functionality of GEIGER running. Roles such as the GEIGER curator and GEIGER support should be clearly defined. The necessary technical measures should be in place to guarantee a sustainable future for the solution after the end of the GEIGER project. This concludes the final phase of validation.

4.3 Flexibility Towards the Future

Having an overall validation plan is necessary and welcome, but we should be aware that the eventual validation activities are likely to deviate from our plan. One area we have surely not yet addressed is how each use case country will have its own implementation of certain validation activities. Depending on the needs of the users and personas in the use case countries, experiments will have to be adapted to address specific rebuttals occurring in these situations. The strength of our argument-based validation framework is that such additions are accommodated for, without having a completely open-ended approach to validation.

Another point to mention is that many of the activities we mentioned overlap with the work done in other work packages. This can be the optimization and hardening work done in T2.5, the educational validations performed in WP3, as well as the surveys and other research of WP5. In each case, we will need to evaluate whether other tasks or work packages take the lead in such validation activities, or WP4 takes the lead. If other work packages take the lead in designing and executing a particular activity related to validation, we must ensure the activity conforms to our standards, both from a research quality perspective and from a data protection perspective.

In any case, as extensive as our plan is, we realize that the future will never turn out as we expect it to be. Therefore, we focused on creating a flexible validation framework and a general validation planning, rather than trying to specify every minute detail beforehand. This offers us flexibility towards the future. Nevertheless, eventually specific activities do have to take place. In the next section we discuss some of these activities in detail.

5 Validation Activities

The work during the first months of our validation work package was not purely centred around theorizing and planning. Month M16 marked the start of our active validation work with the first activities taking place related to validation Phase 1: validation with GEIGER use case partners. We discuss the first validation activities that took place in this section. First, in Section 5.1, we will cover activities spanning all use case countries. Then, in Section 5.2, we will provide details regarding activities or adaptations that were specific to the three use case countries: Switzerland, The Netherlands, and Romania.

5.1 General

Phase 1 of validation constituted several activities spanning all use case countries. One such activity was an expert evaluation of GEIGER content, performed by an expert panel of GEIGER consortium members with different areas of expertise. We additionally initiated two monthly surveys, one among use case partners and one among GEIGER multipliers. Both activities will be discussed in Section 5.1.1.

The first steps in user experience testing were taken in the past months. The process and results are described in Section 5.1.2. As noted, WP4 is heavily linked to the optimisation and hardening task T2.5. In Section 5.1.3 we describe how the two are linked and where we have achieved synergies so far. Lastly, we perform a similar exercise for the relation between the work in WP3 and WP4 in Section 5.1.4.

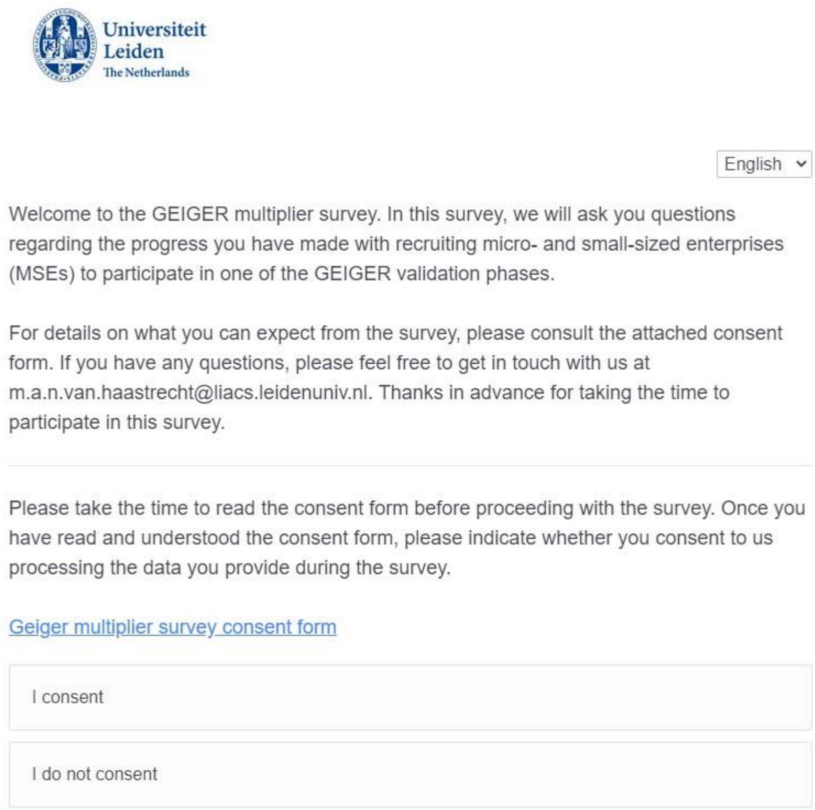
5.1.1 Surveys and Expert Evaluations

Surveys

Phase 1 of validation involved the launch of our first two monthly surveys. The first is the GEIGER multiplier survey, aimed at the GEIGER multipliers SKV (Switzerland), SRA (The Netherlands), and CLUJ IT (Romania). The survey is intended to track the progress of the multipliers in motivating MSEs to become involved in the GEIGER process. In the first phase of validation, this is mainly intended to unearth any problems we are having in attracting users for the second phase of validation (involving alpha users).

Figure 25 shows the welcome screen multiplier partners saw at the start of the survey. We included an extensive consent form based on guidelines provided in deliverable D7.1, to ensure participants were adequately informed of all details pertaining to the survey. The survey was made available both in English and in German.

The second phase of validation requires a minimum of 15 MSEs involved per use case country, to match the diversity in personas we observe. The survey therefore contains questions on how many alpha users have confirmed their intention to participate, how many alpha users are expected to participate in the future, and - especially when too few MSEs have been reached - what help is needed to reach more MSEs. If help is required in a specific use case country, we will ensure together with the partners in WP5, that the necessary measures are taken to reach enough participants.



Welcome to the GEIGER multiplier survey. In this survey, we will ask you questions regarding the progress you have made with recruiting micro- and small-sized enterprises (MSEs) to participate in one of the GEIGER validation phases.

For details on what you can expect from the survey, please consult the attached consent form. If you have any questions, please feel free to get in touch with us at m.a.n.van.haastrecht@liacs.leidenuniv.nl. Thanks in advance for taking the time to participate in this survey.

Please take the time to read the consent form before proceeding with the survey. Once you have read and understood the consent form, please indicate whether you consent to us processing the data you provide during the survey.

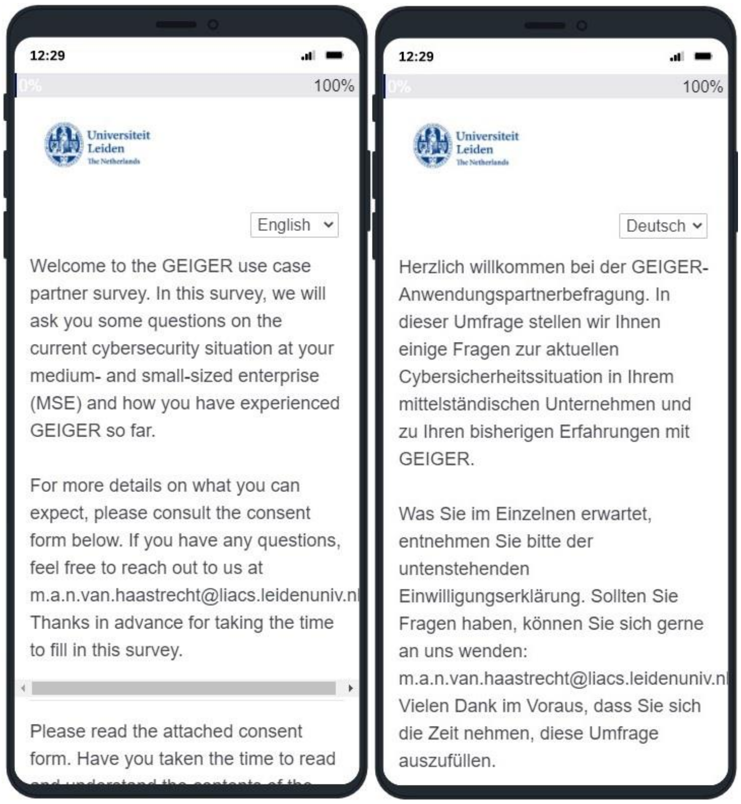
Geiger multiplier survey consent form

I consent

I do not consent

Figure 25: GEIGER multiplier survey welcome page.

The second survey that has commenced is the monthly survey of GEIGER use case partners. The survey is intended to run between months M18 and M29, following the GEIGER use case partners over a long period. The survey will help us to measure the progress made by MSEs during this period, and the effect that GEIGER has had on their cybersecurity awareness and resilience. Figure 26 shows the mobile welcome page for the survey.



Welcome to the GEIGER use case partner survey. In this survey, we will ask you some questions on the current cybersecurity situation at your medium- and small-sized enterprise (MSE) and how you have experienced GEIGER so far.

For more details on what you can expect, please consult the consent form below. If you have any questions, feel free to reach out to us at m.a.n.van.haastrecht@liacs.leidenuniv.nl Thanks in advance for taking the time to fill in this survey.

Please read the attached consent form. Have you taken the time to read

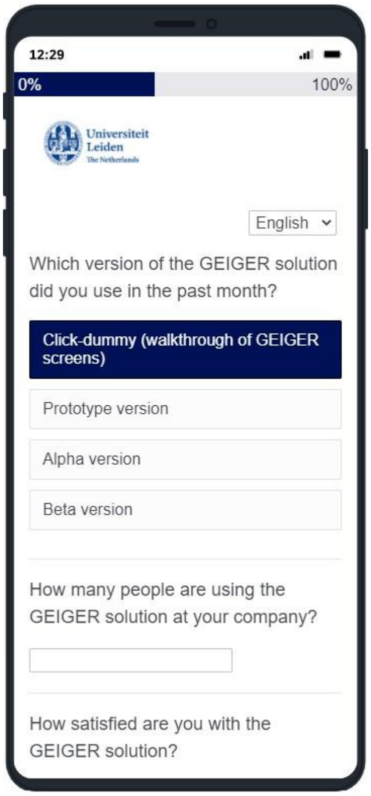
Herzlich willkommen bei der GEIGER-Anwendungspartnerbefragung. In dieser Umfrage stellen wir Ihnen einige Fragen zur aktuellen Cybersicherheitssituation in Ihrem mittelständischen Unternehmen und zu Ihren bisherigen Erfahrungen mit GEIGER.

Was Sie im Einzelnen erwartet, entnehmen Sie bitte der untenstehenden Einwilligungserklärung. Sollten Sie Fragen haben, können Sie sich gerne an uns wenden: m.a.n.van.haastrecht@liacs.leidenuniv.nl Vielen Dank im Voraus, dass Sie sich die Zeit nehmen, diese Umfrage auszufüllen.

Figure 26: The mobile version of our use case partner survey, in English and German.

Again, we made the survey available in both English and German, to cater to the needs of the GEIGER use case partners. We communicated with the Romanian use case partners regarding their language preferences before sending out the survey, but they indicated that English was their preferred language of communication for the project.

Besides containing questions related to the level of awareness and resilience at the MSE, we also include questions related to how satisfied users are with the GEIGER solution and which elements of the GEIGER solution they are using. Together with the key question of whether users have experienced an impactful incident in the past month, the answers to these questions will help us to determine both how satisfied users are with the GEIGER solution and how accurate the GEIGER indicator is in predicting cybersecurity incidents at an MSE. Some example questions are shown in Figure 27.



Which version of the GEIGER solution did you use in the past month?

- Click-dummy (walkthrough of GEIGER screens)
- Prototype version
- Alpha version
- Beta version

How many people are using the GEIGER solution at your company?

How satisfied are you with the GEIGER solution?

Figure 27: Example questions from the use case partner survey.

Expert Evaluations

A key question at the outset of validation was: Where should we focus our attention? Certain inferences and claims regarding the GEIGER solution will have a high level of face validity, meaning rebuttals will be less numerous, and warrants and backings are less necessary. However, there will also be areas where there are many doubts regarding the claims made within the GEIGER project. These are the areas where we need to focus our validation work, to examine the validity of claims in detail and to facilitate an objective judgement regarding the balance between rebuttals and warrants/backings.

To investigate where there may be areas in the GEIGER solution that deserve more attention during validation, we formed an expert panel of partners within the GEIGER project with various areas of expertise, to evaluate the GEIGER content. Table 30 depicts the participants in the GEIGER expert panel and their area of expertise.

Table 30: GEIGER expert panel participants.

Name	Affiliation	Expertise
Rolan Kab	KPMG	Privacy and GDPR
Bernd Remmele	PHF	Education
Amedeo D'Arcangelo	KSP	Security
Moritz Dietsche	HAAKO	MSE
Heini Järvinen	TECH EU	Dissemination
Tony van Oorschot	SRA	Accountants
Bettina Schneider	FHNW	Privacy and GDPR

The participants all received a summary of the GEIGER security, privacy, and GDPR content that is part of the GEIGER education ecosystem and the GEIGER indicator. This summary did not introduce any new knowledge, but rather concisely presented the ideas introduced in earlier deliverables. The participants were asked to read the summary and consult any additional resources such as deliverables where they deemed this necessary. After reading the content, they provided a concise assessment of areas where the completeness and relevance of the GEIGER content could potentially be questioned.

The assessments were used as a starting point for a 2-hour panel discussion to discover potential focus areas for validation. In this session, expert panel participants were able to explain their views on potential pitfalls and areas for improvement. Table 31 presents the main findings resulting from the expert panel session, along with the work packages responsible for providing sufficient argumentation related to the findings.

Table 31: Expert panel findings.

Finding	Responsible WPs
The GEIGER solution should be clear and transparent about which NIST CSF factors it does (identify, protect, detect) and does not (respond, recover) cover, for now. CSDs can help MSEs to respond to and recover from incidents, but they must have a process to follow to do this.	WP2, WP3, WP4
An elegant technical alignment of the GEIGER education ecosystem and the GEIGER indicator, that eases the integration of further educational providers, may not occur during the project lifetime. Clarity with regards to this point should be provided by WP2.	WP2
There must be adequate processes in place, both for the indicator and for education, to adapt to a changing cyber landscape. This is an area where there are currently more question marks than in other areas of GEIGER.	WP2, WP3
There are doubts regarding the MSE classification used. WP4 should evaluate whether the chosen classification is appropriate.	WP4
There are doubts whether the GEIGER indicator adequately addresses web service security. WP4 should take this into account when designing user tests.	WP4
There is some confusion related to the progression through knowledge levels in the educational curriculum. WP3 must evaluate whether these doubts are significant enough to warrant a re-evaluation of the knowledge level progression.	WP3
WP4 and WP5 must collaborate to form a clearer image of the exact GEIGER target audience. This should also help to ensure appropriate messaging, both publicly and within the GEIGER application.	WP4, WP5
There are doubts regarding the certification process for security defenders. Both the specificities of the certification scheme for security defenders (WP3), and the intellectual property regulations regarding (local) training materials (WP5) need to be crystalized.	WP3, WP5
There are doubts about some of the vocabulary currently used within the GEIGER project. We should take care to remove any unnecessary jargon.	WP4, WP5

GEIGER should nudge people towards making a first step towards becoming aware of the intricacies of GDPR. GDPR content should be adequately integrated into the GEIGER application (WP2) and the effects of the GDPR content on users in their context should be investigated (WP4).	WP2, WP4

Moving forward, we will use these results to tailor our validation experiments to ensure that we adequately address these areas of the GEIGER solution. However, this certainly does not imply other areas of the GEIGER solution will receive insufficient attention during validation. The expert panel only serves to identify assumptions in the GEIGER proposal that are worth questioning, not to identify assumptions that are to be left untouched during validation. As we have stressed at length when covering our theoretical validation framework, we are intent on performing rigorous validation in the GEIGER project.

5.1.2 User Experience Testing

IIT (FHNW) conducted a moderated remote usability test using a Figma click dummy on the participant's computer. The click dummy simulated the GEIGER toolbox from the perspective of a first-time user. Furthermore, the download and the use of the tools 'Geiger Mobile Learning', 'Cyberrange' and the plugin 'Device Report' from KSP were simulated.

Phishing and malware were used as representative cyber threats. The download and use of the tools 'Geiger Mobile Learning', 'Cyberrange' and the plug-in 'Device Report' from Kaspersky were simulated as concrete recommendations.

Webex was used to moderate and record the sessions. Each session captured each participant's comments, navigational choices, Task completion rates, questions, and feedback.

5.1.2.1 Research Goals

- **RQ1 How well is the cyber security risk of an SME conveyed to a user?**
  - RQ1.1 (KPI 1.2, CR1.R04.2): How well does the CEO of the MSE understand the GEIGER indicator (5-step Likert question)?
  - RQ1.2 (KPI 1.3): How well do the recommendations match the types of attacks being experienced by the MSE (5-step Likert question)?
  - RQ1.3 (CR1.R05.3): How well does the MSE user understand how and why the GEIGER indicator score has changed (5-step Likert question)?
- **RQ2 Is the user motivated to learn about cyber security?**
- **RQ3 How willing is the user to download external apps?**
  - RQ3.1 (KPI I2.1.3.1, T.QR01.1): How many of the shown top recommendations for education/settings were adopted (number)?[10]
  - RQ3.1 (KPI I2.1.3.2, T.QR01.1): How many of the shown top recommendations for technical protection were adopted (number)?
  - RQ3.3 (KPI I2.1.3.3): How many of the shown incidents were resolved (number)?
- **RQ4 How well does the design of the toolbox work together with the design of external tools?**

[10] For KPI I2.1.3.1-3, we also need to know how many recommendations were shown.

  - RQ4.1 (T.QR06.2): How well do the integrated tools adhere to the GEIGER look-and-feel (5-step Likert question)?
  - RQ4.2 (CR1.R04.2): How easy does the MSE user perceive the use of the recommended tools (5-step Likert question)
  - RQ4.3 (CR1.R05.2): How well does the MSE user understand the information to be entered into the tools (5-step Likert question)?
- **RQ5 How good is the quality of the GEIGER framework as perceived by the SME end users overall?**
  - RQ5.1 (KPI 5.3): How much is the MSE user satisfied with the GEIGER framework overall (5-step Likert question)?
  - RQ5.2 (T.QR06, CR1.R04.2): How easy is it for the MSE user to use the toolbox overall (5-step Likert question, System Usability Scale)?
  - RQ5.3 (CR1.R03.1, CR1.R03.4): Does the MSE user worry about any of the following (O missing functions, O size of the toolbox, O bad performance, O battery consumption, O spamming other apps (mail, calendar, push messages), O how the toolbox works, O bad reliability, O bad security, O difficult to install/update/uninstall, O other)? What exactly (free answer)?
  - RQ5.4 (CR1.R04.2): How easy does the MSE user perceive the installation and configuration of the toolbox (5-step Likert question)?

5.1.2.2 Methodology

The moderator contacted and recruited use case partners from Switzerland, The Netherlands, and Romania as participants. In advance, each participant received an invitation for the Webex room and a consent form to fill out. The sessions lasted approximately between 90 and 120 minutes.

- Test moderator explained the purpose of the session and the scenario of the click dummy and introduced the tasks.

  Clickdummy link:
  https://www.figma.com/proto/EaaV5arK97KRd9PcWpyGHT/Geiger-Toolbox-UI?page-id=1%3A2&node-id=1332%3A9542&viewport=241%2C48%2C0.48&scaling=min-zoom&starting-point-node-id=1041%3A14472&show-proto-sidebar=1

- Specific interview questions were asked after each task. At the end of the last task, the test administrators had each participant fill out a system usability scale and general questions as an online questionnaire.

  Detailed Results with recording timestamps:
  https://drive.switch.ch/index.php/s/ASmw3bu46V0B4Fy

5.1.2.3 Participants

Table 32 All participants of the usability test

Participant	Country	Prior Knowledge
Alain	Switzerland	Knew nothing, has not seen any screen
Loredana	Switzerland (use case partner)	Knew the concept, has seen the indicators
Moritz	Switzerland (use case partner	Knew the concept, has seen the indicators
Heike	Switzerland (use case partner)	Knew the concept, has not seen any screen
Frank	Netherland	Knew the concept, has not seen any screen
Tony	Netherland (use case partner)	Knew the concept, has not seen any screen
Daniel	Romania (use case partner)	Knew the concept, has seen some wireframes
Vlad	Romania (use case partner)	Knew the concept, has not seen any screen

5.1.2.4 Introduction & Pre-Test questions

At the beginning of each session, we gave the participants following instructions and questions:

Introduction
"Thank you for taking the time to participate in our usability test today. The goal of today's session is to test the toolbox of the geiger project. The toolbox is an app that is addressed to small micro enterprises and should help to protect them in relation to cyber security. Today we will give you tasks that you will perform with the toolbox. This is not the final app but a simulation. It is important to know that this test is only about the performance of the toolbox and not about your individual performance."

Pre-test Questions

- What do you already know about the Geiger toolbox?
- Have you ever seen any screen of the toolbox before?

Usability Test Rules

- "During the entire usability test, it is very important to share your thoughts out loud"
- "The session can take up to 120min. Let us know at any time if you want a break."

Scenario Description
"You own a small company with three employees. You want your company to be secure when it comes to cyber security. That's why you downloaded the "geiger toolbox" smartphone app. You are about to open the app for the first time and hope that it will help you protect your company."

5.1.2.5 Tasks

The following tasks were carried out by the participants in exactly this order. At the end of each task, the following post-task questions were asked.

Table 33 Usability test tasks with corresponding questions.

Task	Description	Post Task Questions
1	**Exploring Toolbox** Find out how to use the app and interpret the cybersecurity of your company. Let us know how you interpret each risk value you see.	- What do you have to do in this app? - What's the next step you would take? - What are the differences between 'user risk' and 'device risk'? - What do you think about the current cyber security situation of your SME? - How would the risk indicators behave when you would improve cyber security?
2	**Geiger Mobile Learning** Navigate to phishing and complete the recommendation "Strong passwords". When you are finished, return to the toolbox and interpret your company's cybersecurity situation.	- How do you feel about the toolbox score change? - What do you think of the score from the geiger mobile learning app?
3	**Cyberrange** Navigate to phishing and complete the recommendation "Practice recognizing phishing mails". When you are finished, return to the toolbox and interpret your company's cybersecurity situation.	- What is the purpose of the cyberrange app? - What do you think about the score of the cyberrange app in comparison with the toolbox score? - What do you think about the current cyber security of your SME?
4	**Device Report Plugin** Navigate to malware and complete the recommendation "Activate Device Report". When you are finished, interpret your company's cybersecurity situation.	- What does the device report do? - Would you activate the device report on your real phone?
5	**Add a device to your toolbox** Interpret your company's cybersecurity situation.	- What influence on your geiger risk score does it have when you add other devices?
6	**Add an employee** Add Mitchel Bradbury as an employee to your toolbox. When you are finished, interpret your company's cybersecurity situation.	- What are the benefits for your company when you add an employee to your toolbox?

5.1.2.6 Findings related to all tasks

Scan Risk Button

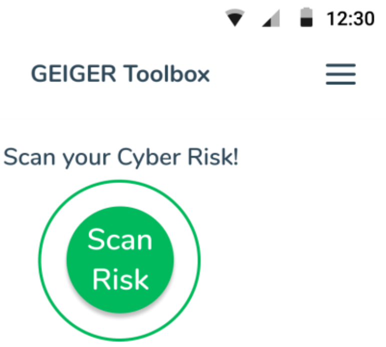
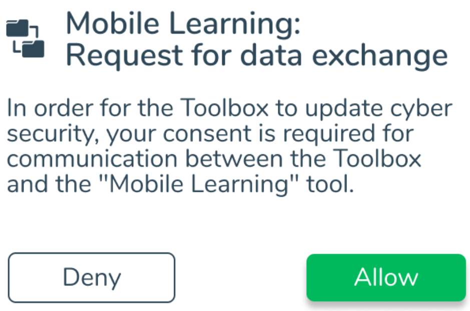


Figure 28 Screen Dashboard: Pulsating Scan Risk Button.

Table 34 Scan Risk Button

Finding	UX Recommendation	State-ments	Impact
For some users it is not immediately clear that the pulsating Scan Risk button should be pressed in order to update the indicator values. Sometimes they thought that the toolbox would already be processing in the background.	The Scan Risk Button must be designed more recognizable in order to signal that the risk indicators can be updated. In addition, while starting the toolbox for the first time, as well as confirming data exchange with new tools, the scan could get triggered automatically.	5	High

Data Exchange Permission



Mobile Learning:
Request for data exchange

In order for the Toolbox to update cyber security, your consent is required for communication between the Toolbox and the "Mobile Learning" tool.

Figure 29 Screen dashboard: data exchange permission dialogue.

Table 35 Trusting the data exchange.

Finding	UX Recommendation	State-ments	Impact
Two users were hesitant to share data with the tools, either they had given their consent at the beginning of the toolbox or it was not clear which data was actually shared. Disclaimer: The confirmation dialogue was only used on 5 participants.	An overall introduction is required about how the toolbox shares data with other tools and why data are handled safely.	2	High

Table 36 Automated scan after permit.

Finding	UX Recommendation	State-ments	Impact
Two users expected that after pressing 'allow' for data exchange that the toolbox would already start processing something, instead of pressing the scan risk button. Disclaimer: The confirmation dialogue was only used on 5 participants.	After confirming a data exchange permission, the risk scan can be triggered automatically.	2	Mediu m

5.1.2.6.1 Indicator related findings

Recognizing Indicator Changes

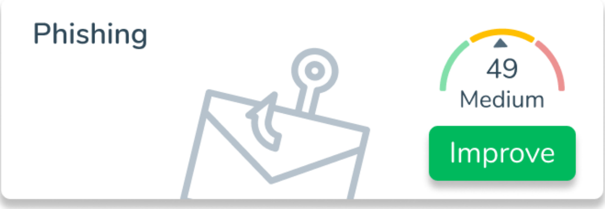


Figure 30 Changed Threat indicator after Task 2.

Table 37 Recognizing Indicator changes

Finding	User Feedback	State-ments	Impact
Most users do not immediately see whether one of the threat risk-, user risk- or device risk indicators has changed. The reason for this could very likely be that the influence on the indicator values was only small.	The effects on the indicators must be made more attention-grabbing with the help of animations. Otherwise the influence on the indicator values should be greater	7	High

Table 38 Unintuitive Improvement Behaviour

Finding	User Feedback	State-ments	Impact
Most users were able to correctly anticipate how the risk indicator would behave in the event of an improvement. Nevertheless, half of the users made similar suggestions how the indicator should behave differently.	Half of the users suggested that if there was an improvement, the indicator value should increase and the pointer should rotate to the right towards a green area on the left side.	4	Medium

Table 39 Unintuitive Improvement behaviour.

Finding	User Feedback	State-ments	Impact

Most users were able to correctly anticipate how the risk indicator would behave in the event of an improvement. Nevertheless, half of the users found the current indicator design to be unintuitive and made some suggestions.	Half of the users suggested that if there was an improvement, the indicator value should increase and the pointer should rotate to the right in the direction of a green area.	4	High

Table 40 Interpretation of Risk Indicators.

Finding	UX Recommendation	State-ments	Impact
After completing a recommendation, most users were able to correctly identify whether the general cyber security of the SME had improved or worsened. One user misinterpreted all indicators throughout the whole session. Another user only interpreted the changes correctly after task 3. The reason here was that the change in the small rotation of the indicator pointer was not noticed. Instead, the indicator value was noted but misinterpreted.	The influence of a completed recommendation should be greater. The radius of the indicator could be larger so that the indicator marker moves more noticeable. In addition, the indicator could be supported with animations. Also an alternative system that is designed to mainly collect points could also be considered.	T2: 6 T3-6: 7	High

Table 41 Indicator changes generate little feeling of reward.

Finding	UX Recommendation	State-ments	Impact
After completing a recommendation, half of the users said immediately updating the risk indicators that the low risk reduction on the total risk or threat risk was quite demotivating. At this point they have not yet noticed the impact on the sub indicators such as user- and device risk. They said the effort and the return do not seem to match.	All progress on any indicator must be somehow immediately recognizable on the dashboard of the toolbox. Alternatively, a recommendation could be introduced as a larger operation with intermediate goals. After completing the whole recommendation the risk reduction in the toolbox would be greater.	4	High

Table 42 Differentiating the meanings of different risk indicators.

Finding	UX Recommendation	State-ments	Impact
The different risk indicators like 'threat risk-', 'device risk-', 'user risk-' and 'total device risk-' look so similar that some users think they represent the same kind of risk instead of different ones. Those users were only able to differentiate them after completing task 3-4 as the values began to differ more widely. This led to confusion for the users and some had the feeling that this was rather a bug in the test	A reduction in the number of indicators with different meanings would ease the complexity in communicating different risk values. Otherwise the individual indicators need to be introduced and visually distinguishable for each kind of indicator category.	4	High

instead of an intention by the toolbox.			

Inconsistent Indicator Designs

After each task in which a tool was involved, the question was asked how well the respective indicator was perceived. The following could be observed over the entire user experience:

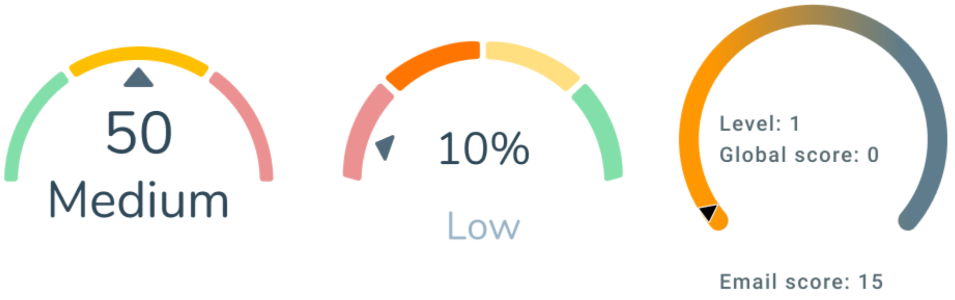


Figure 31 Toolbox indicator, Geiger mobile learning indicator, cyberrange indicator.

Table 43 Inconsistent Indicator Designs.

Finding	UX Recommendation	State-ments	Impact
Users are confused by the number of different indicator designs that work according to different principles. Three users already mentioned that issue after Task 2 and 4 additional users mentioned this issue after Task 3. This could be because the geiger mobile learning is visually more closely related to the toolbox indicators and the inconsistency became more obvious after seeing another indicator from the cyberrange tool. On the one hand, it leads to frustration that once an indicator-/scoring system has been learned that there is a new one for each additional tool. On the other hand, it is difficult to understand what influence a change from a tool indicator has on the toolbox.	In order for the interaction between toolbox and tools to work, a uniform, superordinate scoring / risk indicator system is required which is uniform in terms of values, behaviour and design. Users recommended that the toolbox and tools should share the same rating system so that the influence of a recommendation can be better understood. In terms of design, the indicators / reward points should be designed uniformly so that once a system has been understood, the user can immediately understand it anywhere and immediately.	7	High

Table 44 Geiger Mobile Learning Indicator.

Finding	State-ments	Impact
The majority of users could recognize a change of the learning score from 0 to 10% immediately. Users interpret it correctly that they still have to do some further lessons. - One user thought when one lesson is 10% then it cannot be that long to complete everything. - The motivation why the user should improve could be greater.	6	Medium

- One user associated the change even as a negative reward - One user mentioned that It is not very obvious how the learning score is composed through the learning results		

Table 45 Cyberrange Learn Indicator.

Finding	State-ments	Impact
The majority of users could recognize a change of the cyberrange score but it was not immediately understandable what exactly the cyberrange indicator says. - E.g. "Does not understand “Global Score'"	6	High

Table 46 Inconsistent Total Risk Indicator Behaviour.

Finding	UX Recommendation	State-ments	Impact
Most users expect that adding a device with a risk value of 50 should have a negative effect as the risk value is actually higher than their current total risk. In fact, device values of 50 have no effect. Their conclusion is the result of previous experiences with the toolbox indicator. Therefore this event resembles an inconsistent indicator behaviour. Most could anticipate correctly that higher scores would worsen the security and vice versa.	The total device risk indicator should behave consistently. Otherwise an explicit explanation is required why the device score of 50 has no effect.	8	High

Table 47 Interpreting employees Total Risk Score.

Finding	UX Recommendation	State-ments	Impact
Half of the users said that they do not really understand what the total risk represents and how it is composed. As a result, the total risks for employees are not fully understood as well.	Either a simpler indicator system is needed with a smaller number of indicators in order to intuitively convey the influences. Otherwise, the dependencies of the various indicators must be better introduced at the beginning and changes communicated with additional animations.	4	High

5.1.2.6.2 External Tools

Table 48 Weak Motivation Downloading Additional Apps.

Finding	UX Recommendation	State-ments	Impact
Most users are rather weakly motivated to download several mobile apps and prefer a	Users suggested integrating all tools in one app. The smartphone could generally represent a rather unfavourable platform choice for a brokerage service	6	High

solution where everything would be integrated into one app.	for other tools. From the user's point of view, a web application would have the advantage that the user would not have to load the device directly with tools.		

Table 49 Transition between toolbox and tools.

Finding	**UX Recommendation**	**State-ments**	**Impact**
It was positive that after completing a single lesson, a total of 3 users wanted to complete further lessons on their own initiative. Nevertheless, for the majority it was rather unclear what should be done next immediately after completing the lesson. Some users had to reread the task text of the test to continue, a user really relied on the help of the moderator. It was positive that after completing a single lesson, a total of 3 users wanted to complete further lessons on their own initiative. Nevertheless, for the majority it was rather unclear what should be done next immediately after completing the lesson. Some users had to reread the task text of the test to continue, a user really relied on the help of the moderator. Since the test did not take place on a real smartphone, this could also be a reason why the users did not switch back to the toolbox on their own.	The tools should clearly communicate what users achieved and when they completed a recommendation from the toolbox.	8	Medium

Table 50 Distracted User Experience.

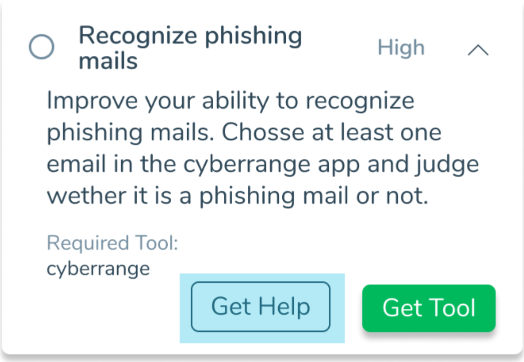
Finding	**UX Recommendation**	**State-ments**	**Impact**
Two users got distracted by visiting the settings of the geiger mobile learning tool during task 2, instead of continuing with the recommendation. One user mentioned that it is not entirely clear which app is now the central administration app.	In every tool, it should be always clear what the next goals are so that the user does not lose track. These goals must be in line with the toolbox recommendations.	2	Medium

Table 51 Button 'Get Tool'.

Finding	**UX Recommendation**	**State-ments**	**Impact**
For one user it was not immediately clear what to expect of the button 'Get Tool'.	Name of the Button could be changed to "download tool".	1	Low

5.1.2.6.3 Security Defenders

Access To Security Defenders



○ Recognize phishing mails High

Improve your ability to recognize phishing mails. Chosse at least one email in the cyberrange app and judge wether it is a phishing mail or not.

Required Tool:
cyberrange

Figure 32 Access to security defender via recommendation card.

Table 52 Access to Security Defenders.

Finding	UX Recommendation	State-ments	Impact
Half of the users commented that the UI might be difficult to find help from a security defender.	More direct navigation to security defenders would be necessary so that this feature is found at the right moment.	4	High

Understanding The Purpose

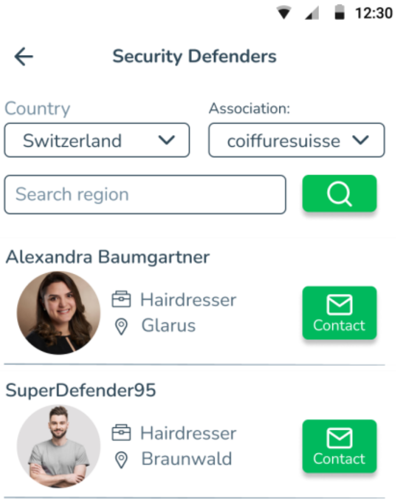


Figure 33 Screen: Security Defenders.

Table 53 Understanding the purpose.

Finding	UX Recommendation	State-ments	Impact
Half of the users mentioned on seeing the security defender screen that it is not automatically clear what	The concept of security defenders should be	4	High

security defenders are good for and when they are supposed to be contacted.	introduced at an appropriate time.		

5.1.2.6.4 Recommendations

This chapter is primarily aimed at the 'Recommendation' screen, which is called up when 'Improve' is clicked in the event of a cyber threat.

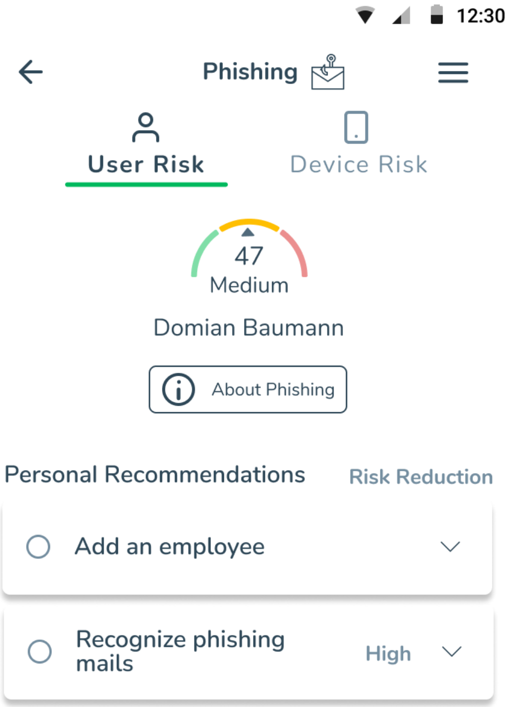


Figure 34 Screen: Recommendations.

Understanding different Recommendation Categories

Figure 35 Recommendation Categories.

Table 54 Understanding different Recommendation Categories.

Finding	**User Feedback**	**State-ments**	**Impact**
Five users could correctly guess what the purpose is of user- and device recommendation. With three users, the interpretation was only partially correct.	A few users who correctly interpreted the purpose of the two categories nevertheless suggested that the concept should be better introduced. Suggestions from users: - to improve the wording of the tabs - to explain what 'user risk' and 'device risk' is	8	Medium

Recognizing ‘Device Risk’ Tab

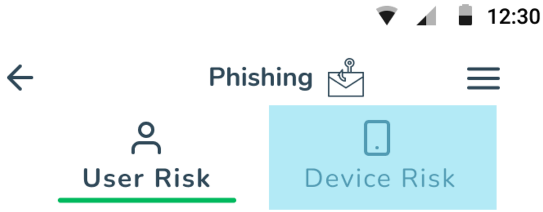


Figure 36 Device Risk Tab.

Table 55 Recognizing ‘Device Risk’ Tab.

Finding	UX Recommendation	State-ments	Impact
One user did not immediately recognize the tab 'Device Risk' and another one could not find it at all.	The 'device risk' tab could either be made more obvious or merged on the same screen with user risk.	2	High

Risk Reduction Label

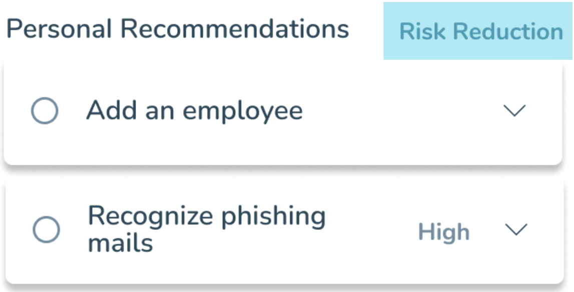


Figure 37 Risk Reduction Label.

Table 56 Risk Reduction Label.

Finding	UX Recommendation	State-ments	Impact
Two users thought the label for the risk reduction level was an interactive element.	Risk reduction label needs to be redesigned to not mislead users.	3	Medium

Risk Reduction Level

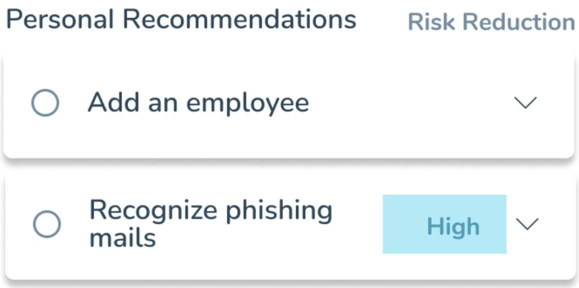
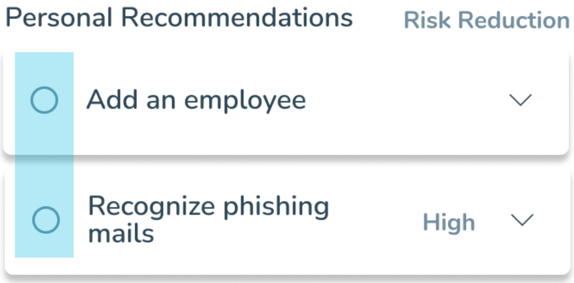


Figure 38 Risk Reduction Level.

Table 57 Risk Reduction Level.

Finding	UX Recommendation	State-ments	Impact
Three users mentioned that they were not able to understand what the purpose of the risk reduction level on a recommendation was.	The Risk Reduction Level must be designed more clearly and must communicate more transparently about the effort & benefit of doing a recommendation.	3	High

Misleading checkmark placeholder


Personal Recommendations

Risk Reduction

○ Add an employee

○ Recognize phishing mails High

Figure 39 Checkmark placeholder.

Table 58 Misleading checkmark placeholder.

Finding	UX Recommendation	State-ments	Impact
Clicked on the left side of a recommendation instead of the arrow on the right side.	Remove the checkmark placeholder on the right side of the recommendation card and only display the checkmark when the recommendation is done. Further to expand the recommendation card, let the whole card be interactable and not only on the small arrow icon.	6	Medium

Table 59 Benefit of Learning/Training.

Finding	UX Recommendation	State-ments	Impact
The recommendations should be designed in such a way that the plausible security increases. For example, it is questionable how effective the sole theory of passwords is. If the security only improves with the effective change of existing passwords, the entire lesson / tutorial should be designed to implement this in the end.	"Recommendations must be designed in such a way that their fulfilment has a clearly traceable safety-relevant benefit. Further it should be explained when and how simply learning/training could already improve cyber security."	2	Medium

Table 60 Consistent Naming.

Finding	UX Recommendation	State-ments	Impact
One user noticed a difference between the labelling of the recommendation title and the mobile learning lesson, which led to confusion.	Recommendation labels and titles of tool lessons should be the same.	1	Medium

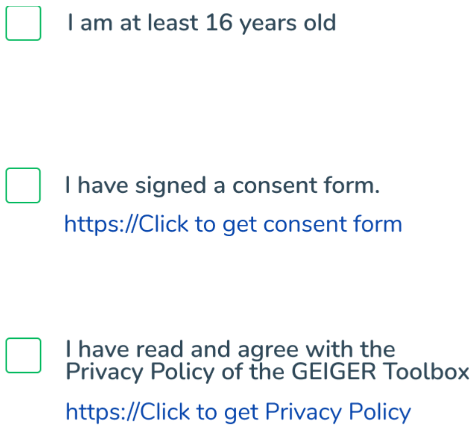
Table 61 Understanding Threat and Recommendation Structure.

Finding	UX Recommendation	State-ments	Impact
Two users mentioned that it is not immediately clear that the recommendation screen belongs to a specific cyber threat as a sub-chapter.	A more present cyber threat icon and name could clarify the affiliation of the recommendation.	2	Low

5.1.2.7 Task 1 Exploring Toolbox

Task 1 was about starting the toolbox as a first-time user to get an idea of the current situation and the existing functions. For this task, navigation was limited to the toolbox. Downloading tools to do recommendations has been blocked and explained to the user.

Terms & Conditions Integration



□ I am at least 16 years old

□ I have signed a consent form.
https://Click to get consent form

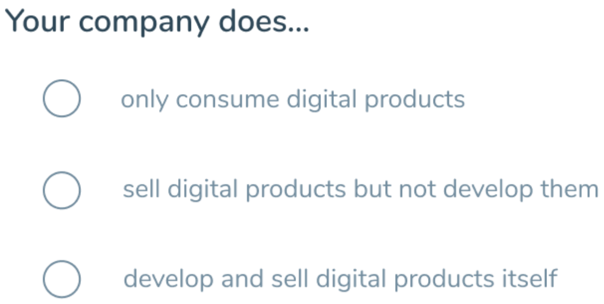
□ I have read and agree with the Privacy Policy of the GEIGER Toolbox
https://Click to get Privacy Policy

Figure 40 Terms & Conditions with corresponding links.

Table 62 Terms & Conditions Integration.

Finding	UX Recommendation	State-ments	Impact
Users just click through the terms & conditions without really reading anything.	Users prefer to view and sign the consent form & privacy policy within the toolbox.	2	Low

SME Category Selection



Your company does...

○ only consume digital products

○ sell digital products but not develop them

○ develop and sell digital products itself

Figure 41 SME Category Selection.

Table 63 SME Category Selection.

Finding	UX Recommendation	State-ments	Impact
Users sometimes struggle to understand the differences between the three SME categories they have to choose.	In order to better distinguish the SME categories, clearer descriptions are needed for each category.	4	High

Table 64 Missing Required Onboarding Questions.

Finding	UX Recommendation	State-ments	Impact
One user rightly asked how the toolbox knows the name of the company and the user.	The current onboarding walkthrough needs to be completed to get all required information for SME owners as well as for employees. The required information could be defined with a simple low-fi wireframe. At this point all technical stakeholders need to verify whether something is still missing or not. In a second phase, the wireframes should be designed in a hi-fi version as the basis for the front-end development.	1	High

Table 65 Recognizing inactive buttons.

Finding	UX Recommendation	State-ments	Impact
One user was unsure if the button 'continue' may be pressed because it was not obvious that the button changed from an inactive into an active state after all terms & condition checkboxes had been filled out.	Redesign inactive state of buttons.	1	Low

Unclear cyber situation of the company

After an initial cyber risk scan, the toolbox lists all current cyber threats with a risk value of 50 for each indicator regardless of the real cyber risk of a company. This leads to two main problems:

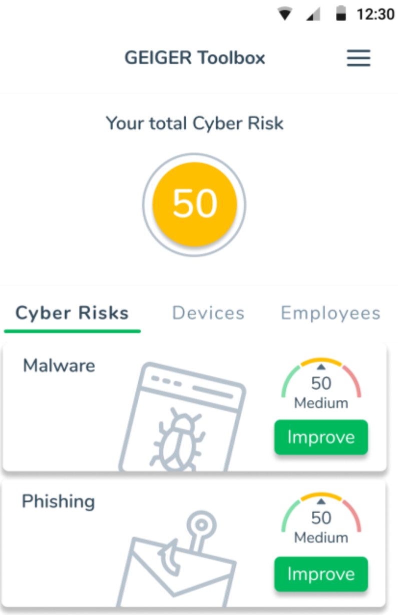


Figure 42 Screen: Dashboard after initial start of the toolbox.

Table 66 Unclear cyber situation of the company.

Finding	UX Recommendation	State-ments	Impact
Almost all users take the initial value of 50 as a real evaluation indicator instead of a neutral starting value. This assumption gives rise to misunderstandings. Technically savvy users desired a relevant statement otherwise it should not be shown at all.	The cyber risk values should be plausible and communicate more transparently what could have been evaluated and what not. As long as the values can not be really evaluated, no misleading placeholder values should be presented. An alternative system of indicators could then be more suitable. For example one that is designed exclusively to collect points for any kind of improvement.	6	High

Table 67 Interview Question: What do you think about the current cyber security situation of your SME?"

Answers	State-ments
“It appears fishy that all the initial risk scores are set to 50”	1
“Thinks the cyber risk of the company is Medium, not really bad but also not really good instead of interpreting that the toolbox only shows current cyber threats but the value has been evaluated yet.	3
“The security is rather low”	2
“The current risk values do not really say anything useful”	2

Table 68 General Understanding of the Toolbox.

Finding	Users Recommendations	Impact
The majority of the participants were able to roughly anticipate what can be done in the toolbox. Nevertheless, throughout the whole test there were several individual comments to suggest how to better convey the purpose of the toolbox:	- An introduction is required in order to gain trust in the toolbox. For example it should be explained how the toolbox protects the company from revealing sensitive data. - An explanation how the toolbox is supposed to be a mediator between cyber tools. - Clear Roadmaps with milestones to show a path of possible achievements. - The toolbox should overall tell clearer what needs to be done next.	Medium

Table 69 Interview Question: "What is the purpose of the toolbox?"

Answers	State-ments
Good understanding: - "Does know that one could improve different cyber security topics, and devices and employees could be added"	6
Partial understanding: - "Should renew their passwords" - "Support of learning is one of the main purpose of the toolbox"	2

Table 70 Introduction of Cyber Risks.

Finding	User Feedback	State-ments	Impact
Basically, users like that they can receive an introduction for each cyber threat and that the information is split bit by bit using illustrations and text.	- The number of slides for intro- ducing a cyber topic should be reduced to a maximum of 4-5 slides. - Clicking and reading through 'About Malware / Phishing' should be rewarded.	5	Low

Table 71 Knowing what to do next.

Finding	UX Recommendation	State-ments	Impact
The value of 50 leads to the fact that the user has difficulties to prioritize which cyber threat or which recommendation should be done next.	Either there is an assessment of the cyber risk of a company right from the start, which is informative and generates values that allow the cyber dangers to be graded into a clear prioritization. If this is not possible, the information architecture of the toolbox should be redesigned from scratch.	6	High

Table 72 Interview Question: What's the next step you would take?

Answers	State-ments
"Would first try to find out what is wrong and then resolve the issues"	2
"Would do what appears to be the most important in the hierarchy and promises a high level of benefit"	1
"Would try to improve malware" (First cyber threat)	1
"Would be curious to see what one can do on 'devices'"	1

5.1.2.8 Task 2 Geiger Mobile Learning

Task 2 was about downloading the tool 'Geiger Mobile Learning' and to complete the lesson 'Strong Passwords' to improve the risk of phishing.

Table 73 Finding the right lesson.

Finding	UX Recommendation	State-ments	Impact
One User was unsure how to navigate to start the lesson because he couldn't exactly remember the description of the toolbox recommendation.	Tools should allow access to the start of the recommendation as directly as possible.	1	High

Table 74 Navigation and Scrolling.

Finding	UX Recommendation	State-ments	Impact
One user got stuck during the password lesson because he could not see that more content could be displayed by scrolling.	It should be made clear at all times that even more content can be displayed using scrolling. The navigation arrows can also only be displayed at the very end so that scrolling has higher priority so that no content is overlooked.	1	Medium

Table 75 Quality of lesson content.

Finding	UX Recommendation	State-ments	Impact
Three users mentioned that some content of the password lesson might be outdated.	The quality of the learning content must correspond to the latest findings in cyber security.	3	High

Table 76 Quiz Result Presentation.

Finding	UX Recommendation	State-ments	Impact
Two users mentioned that on the quizz results it is difficult to distinguish which answers were right or wrong and which would be the correct ones.	The colors of the results should be improved.	2	High

Table 77 Learning Reminder.

Finding	UX Recommendation	State-ments	Impact
There are very different comments on the reminder function of the geiger mobile learning app	The reminder function should be adaptable so that the user has control over the intrusiveness of reminders.	3	Medium
User Feedback - "Would like to receive messages and to have a reminder entered in his personal calendar" - "Would only set a reminder if they don't fill his personal calendar" - "Suggests that the reminder would be already set and that the user has the option to opt out"			

5.1.2.9 Task 3 Cyberrange

Task 3 was about downloading the tool 'Cyberrange' and to evaluate at least one email to improve the risk of phishing.

Table 78 Missing Introduction.

Finding	UX Recommendation	State-ments	Impact
Most of the users could answer the question correctly what the purpose of cyberrange was. However, half of the users did initially not know the purpose of cyberrange and desire a short introduction.	The app needs to have an introduction about the purpose of the cyber game.	4	High

Table 79 Too Real Looking Fake Emails.

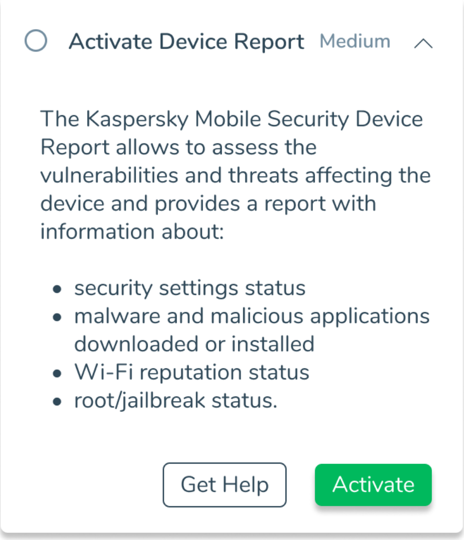
Finding	UX Recommendation	State-ments	Impact
The fake phishing mails from cyberrange look quite realistic. Because of that two users hesitated to press anything since it might cause trouble.	The realism of the fake email should either be weakened or it should be explained that it is only simulated fake emails.	2	Medium

Table 80 Result Overview.

Finding	UX Recommendation	State-ments	Impact
Two users mention that they would like to better understand the results after judging a fake email.	Results should be displayed more transparently to distinguish what was right or wrong.	2	Medium

5.1.2.10 Task 4 Device Report Plugin

Task 4 was about to activate the plugin 'Device Report' to improve the risk of malware.



○ Activate Device Report Medium

The Kaspersky Mobile Security Device Report allows to assess the vulnerabilities and threats affecting the device and provides a report with information about:

- security settings status
- malware and malicious applications downloaded or installed
- Wi-Fi reputation status
- root/jailbreak status.

Figure 43 Device Report Plugin.

Table 81 Bad Understanding of the plugin.

Finding	UX Recommendation	State-ments	Impact
After activating the device report, most users hardly understand what it is actually doing.	In order to build trust in the toolbox, plugin activity should be communicated transparently. Such activities could be conveyed with the help of illustrations and animations.	7	High

Table 82 Trust Issues.

Finding	UX Recommendation	State-ments	Impact
Generally users liked the little effort it took to improve the risk, but half of the participants have trust issues with the plugin because the current textual information does not sufficiently describe what the plugin actually is doing.	Descriptions about plugins could be more transparent and better conveyed with the help of illustrations and animations.	4	High

Table 83 Recognizing Plugin Activation.

Finding	UX Recommendation	State-ments	Impact
One user did not recognize at all that the plugin had been activated. Three users desired more feedback right after the activation.	Users commented that they would appreciate either to see right away how such a report would look like or to generally receive more feedback regarding what exactly is going on in the background.	1	Medium

5.1.2.11 Task 5 Adding a Device

Task 5 was about to add another android phone.

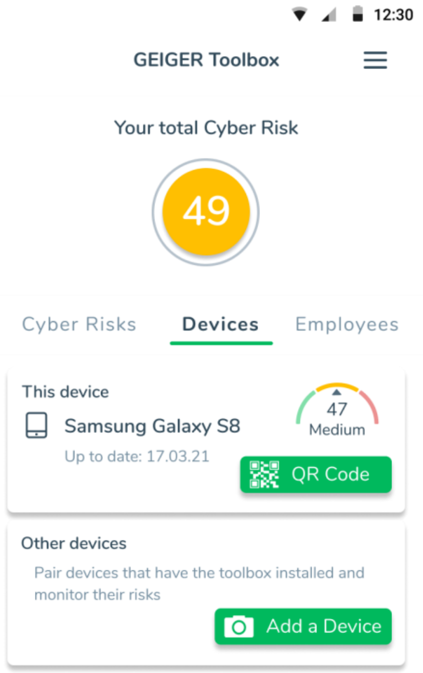


Figure 44 Screen: Devices

Table 84 Recognizing Added Device.

Finding	UX Recommendation	State-ments	Impact
For two users it was not immediately clear that adding a device with the qr code worked and was effectively added to the toolbox.	The user could receive more feedback when a device has been successfully added.	2	High

Table 85 Unclear Influence of Adding Devices.

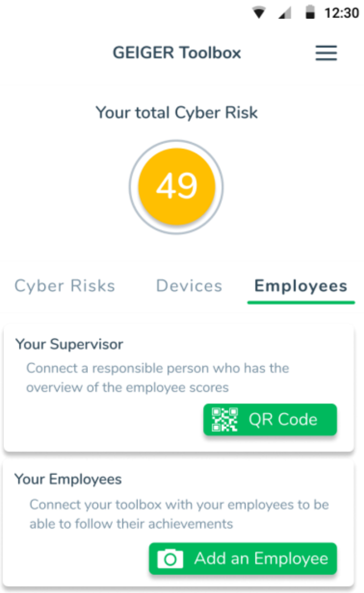
Finding	UX Recommendation	State-ments	Impact
Some users desire more information about why they even should add a device.	An illustrative explanation is required of what the benefits of adding devices are.	2	High

Table 86 Unclear Influence of Adding Devices.

Finding	UX Recommendation	State-ments	Impact
Some users have trouble distinguishing the influence of user and device-specific measures, they think that the new device should already have implemented measures recommendations like the 'Strong Password' lesson.	An illustrative explanation is required which actions influence the risk of a device.	3	High

5.1.2.12 Task 6 Adding an Employee

Task 6 was about to add an employee called 'Mitchel Bradburry' and to request his total risk indicator.



Connect a responsible person who has the overview of the employee scores

Connect your toolbox with your employees to be able to follow their achievements

Figure 45 Screen: Employees

Table 87 How to Approach Employees.

Finding	UX Recommendation	State-ments	Impact
Most users know which employee to approach but they do not know fully the concept of being a supervisor in the toolbox and how they should deal with employees with high risk values. One reason is that the employee's total risk does not tell them much whereas one user liked the data privacy of an unspecific risk indicator	Users need an introduction about the role of supervisors within the toolbox and why an unspecified score makes sense for reasons of privacy protection, but also how they can specifically support their employees.	6	High

Table 88 Negative Feelings Towards Employees.

Finding	UX Recommendation	State-ments	Impact
Two users responded with the attitude that the employee did a poor job when they saw the higher employee indicator. Depending on how a supervisor interprets the employee indicator, this could lead to a negative experience for employees.	An indicator system based on collecting points could avoid the impression that employees pose a threat to a company and rather could benefit from support.	2	High

Table 89 Missing Request To Access Camera.

Finding	UX Recommendation	State-ments	Impact
The Current Clickdummy does did not involve requesting permission for accessing the	Permission to use the camera should only appear if the camera has to be used in the	1	High

device	appropriate place.		

5.1.2.13 Post Test Questionnaire

After the usability test session all participants were given a link to an online questionnaire. Most questions are those asked by the EU.

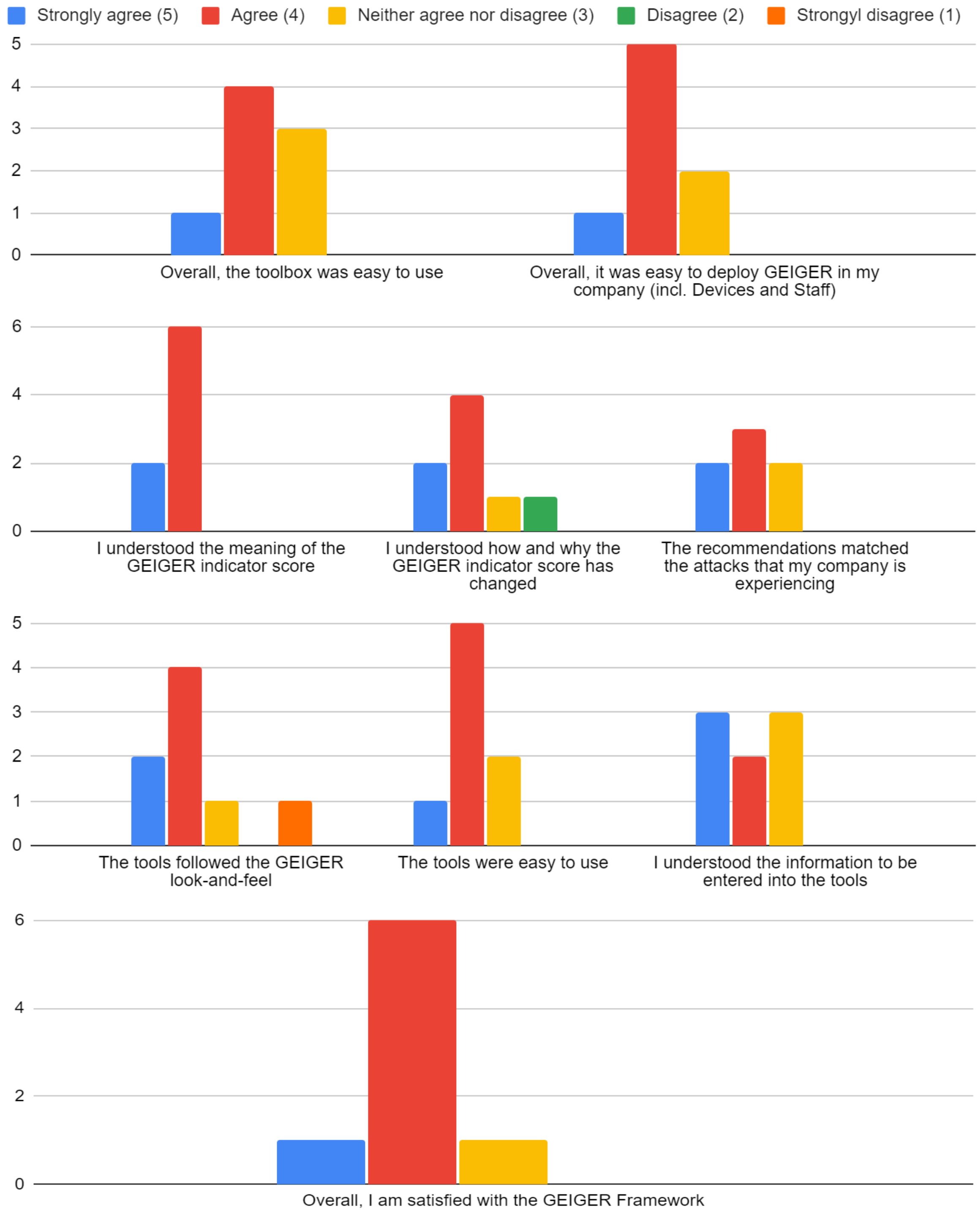


Figure 46 Results of the post-test questionnaire part 1.

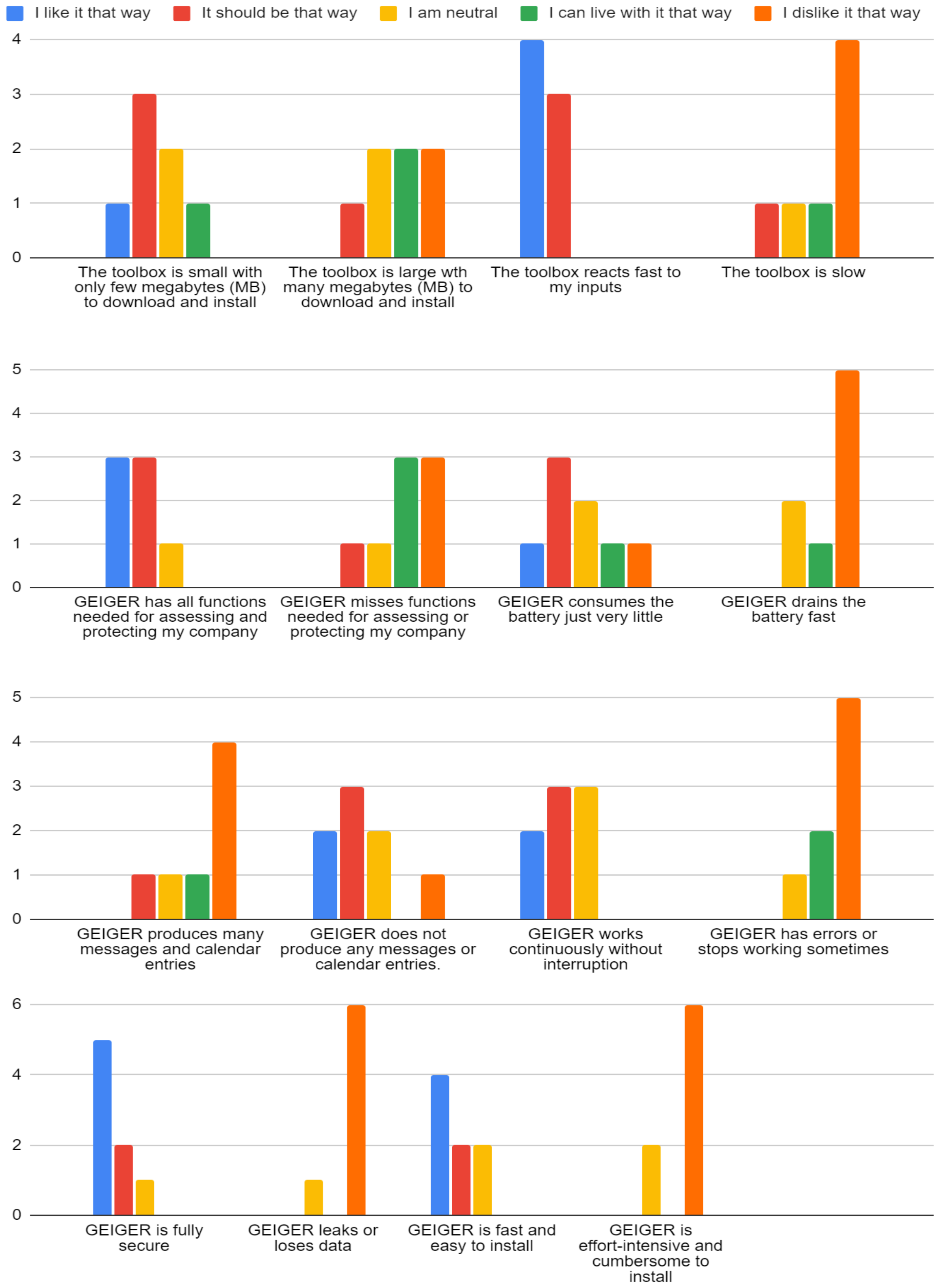


Figure 47 Results of the post-test questionnaire part 2.

5.1.3 Technical Evaluations

5.1.3.1 Continuous Integration Environment to Automate Testing

In the context of WP2 (T2.4 and T2.5), the testing team set up a test a CI/CD environment to automate the integration between different GEIGER solution modules and the delivery of an integrated solution allowing to perform different kinds of functional and non-functional tests including Unit, Integration, Performance and Security testing.

The main objective of this environment is to facilitate the technical evaluation of the GEIGER Solution in terms of distinct elements/modules and their potential systemic interaction. The list of modules can be summarized in the following:

- The GEIGER toolbox
    - Local knowledge base including scanning metrics and recommendations
        - KSP SDK
        - MI Cyberrange
        - CySec
        - KPMG Chatbot
        - Risk indicator computation
        - Recommendations
    - GEIGER Indicator
- Geiger Cloud
    - Gather information of external and infrastructure apps
        - Kaspersky Interactive Protection Simulation (KIPS)
        - CSMG
        - Montimage IDS
    - Information provided by CERTs and CSIRTs
    - Information synchronized with the GEIGER Toolbox.
- Device pairing

The results of these technical evaluations will be reported in WP2 including hardening solutions for security testing.

5.1.3.2 Helpdesk and MSE Support During Piloting

Even the GEIGER solution tested before starting the piloting trails with third-party MSEs, we believe that technical support is needed. For this purpose, we defined three levels of support.

The first level of support will be detailed documentation accessible on the GEIGER webpage, including a Frequently Asked Questions (FAQ) page.

The second level of support will be provided by the MSEs that are members of the GEIGER consortium. Thanks to their involvement in earlier trials (M13-M24), they will have experience using GEIGER in an MSE and can offer advice and guidance to third party MSEs. This support will be provided in webinars advertised to the MSE target audience and during the onboarding of MSEs into the trials. As part of these webinars, also security defenders experienced in helping to secure an MSE will be encouraged to participate and share lessons learned.

The third level of support will be interactive assistance to answer questions concerning the expectations on the MSE within the trial, the technical functioning of the GEIGER solution, and data protection. These requests will be answered by WP4 partners (questions concerning involvement in the pilots), T2.4/T2.5 partners (questions concerning the technical functioning of GEIGER), and WP4/WP7 (questions concerning data protection). A schedule will be established to manage the staffing and availability of this third-level support.

Technical support has been foreseen in Task 2.5 and will be performed mainly by Montimage and tools providers, trial support in WP4 with one use case leader in each use case country (Switzerland, The Netherlands, Romania). The onboarding of the MSE participants in a trial, including how consent is obtained and managed, has been defined in D7.1. The data protection procedures and responsibilities have been defined in D7.2.

5.1.4 Education Evaluations

Validation of the GEIGER educational ecosystem relates to the distinct elements and their potential systemic interaction.

The main elements that have been discussed above are:

- Exploiting specific learning/training methodologies: experiential learning, particularly game-based learning, and reverse mentoring.
- Partly in relation to that the specific educational aspects of the different use cases, e.g., reverse mentoring of the Swiss (non-IT) apprentices or the experiential approach to Romanian entrepreneurs.
- The effectiveness of the GEIGER training features to support the general learning goals – content wise and in concern of accessibility, e.g., MI cyberrange to improve Cybersecurity awareness about phishing threat, or the availability of Mobile Learning by the CySec Adaptation as well as the inclusion of GEIGER-Related Topics (how to do, how to communicate).
- The implementation of effective Train-the-Trainer-Scheme.
- The feasibility (or even effectiveness) of the two-sided Interoperability of the GEIGER-Curriculum, i.e., curricular interoperability with further educational providers regarding course development etc. and the technical interoperability with the toolbox (also for additional contributors).
- Partly in relation to that the liveliness of the GEIGER-Communities.

5.2 Use Case Countries

Although certain validation activities apply to all GEIGER use case countries equally, most will require at least some adaptation to the specifics of each situation. In this section we will cover some of these adaptations, as well as introducing the personas involved in each use case country. The different personas mandate a different validation approach. On the one hand this poses difficulties for the validation process, but as we will see in the following sections, it also adds depth to our validation procedure that we would otherwise not be able to achieve.

5.2.1 T4.1: Switzerland

The Swiss use case consists of the following personas.

- Apprentices (non-IT) [**ANI**]
- Apprentices (IT) [**AI**]
- Teacher (non-IT) [**TNI**]
- Teacher (IT) [**TI**]
- Students (non-apprenticeship) [**S**]
- Cybersecurity experts [**E**]
- MSE-Manager (digitally dependent [**DD**], digitally based [**DB**], digital enabler [**DE**])
- Associations [**ASC**]
- Educators [**ED**]

- Tool Vendors [**TV**]

In preparation for the validation phase, we looked at the requirements of the personas for the training. This resulted in the journey shown in Table 90.

Table 90: Journey in the Swiss use case.

Activity	**Persona**	**Period**	**Description**
Awareness	All	M12 – M30	Creating awareness. Informing
Onboarding	All	M17 – M30	ANI and AI are boarded through their TNI and TI. After this first round, onboarding will be conducted through events during the remainder of the project duration.
Education / Training	TI, NTI ANI, NI, S DD, DB, DE	M14 – M27 M17 – M30 M22 – M30	Education and training start with TI and NTI, who subsequently train their classes. Full education and certification will be organized in several small groups and different courses to meet the needs of the different ANI, NI and S. DD, DB and DE start training after a first event in March 2022.
Validation / Pilot	TI, NTI ANI, NI, S DD, DB, DE	M14 – M27 M17 – M30 M22 – M30	Validation is executed according to the argumentation of the personas and after each training sequence. This approach allows for improvement in training or training materials before the use in the next small group.
Evaluation	TI, NTI ANI, NI, S, DD, DB, DE	M18 – M30	Evaluation is done after concluding larger training and validation sequences such as L1, L2, L3.

To be able to uncouple education and training from the availability of tools and learning materials as much as possible and to meet the needs of our different TI, TNI and S we are choosing an individualised approach. After the start of education and training with the train the trainer session and certification in the CSMG-Game, the certified teachers subsequently adapted the pedagogical approach as needed and applied the first general cyber security training in their classes as awareness raising. Interested apprentices are now enrolled to attend the full training. Feedback agreed on with WP3 has been carried out. Validation will be done after having developed a suitable instrument in WP4.

Using the experience gained from this first part of training and education, the planning of the whole Swiss use case has been reviewed, adapted, and complemented in the EduHack workshop on the 15th and 16th of November 2021 in Baden.

At the EduHack we created an overview of the currently planned initiatives and epics. It is important to note, that the whole validation and evaluation journey hasn't a linear course and connects many personas as well as all other WP partners in a way or other. Therefore, we chose to do the planning in an agile way with epics and initiatives and go on to refine them to user stories for the actual execution.

On initiative level, the Swiss use case currently has the following cards:

- Develop education content
- Conduct education
- Support building the community
- Local launch event CH
- Involve SME&ME and associations

- Define Swiss certification procedure CSDs
- Test GEIGER app with use case partners and security defenders
- Content for pitches
- Build educator community
- Validation and Dissemination

5.2.2 T4.2: The Netherlands

5.2.2.1 Preparation Phase

In addition to the definition of the use case requirements analysis and solution specifications for the Dutch use case (WP1, task 1.1) several provisionary personas were defined during a brainstorm meeting[11]:

1. **TA-K** : **T**rusted **A**dvisor (i.e., optional), **K**nowledgeable, external.
2. **TA-U** : **T**rusted **A**dvisor (i.e., optional), **U**ninformed, external.
3. **CA-K** : **C**hartered **A**ccountant (i.e., compulsory), **K**nowledgeable, external.
4. **CA-U** : **C**hartered **A**ccountant (i.e., compulsory), **U**ninformed, external.
5. **ITM** : **IT M**anager (i.e., optional), knowledgeable, internal.

Looking more closely at these personas we concluded that the focus of the accountant as a trusted advisor is based on business risks rather than financial risks. The role of trusted advisor applies to both type of accountants in The Netherlands:

- **AA-**accountant : Accounting Consultants (Dutch: **A**ccountant **A**dministratie consulent)[12]
- **RA-** accountant : Chartered Accountant (Dutch: **R**egister **A**ccountant)[13]

These two types of accountants can be supported by IT-auditors[14] (with title RE / CISA) in providing impartial assessment and advice on the quality aspects of IT. Some accountants also are qualified as an IT-auditor. Usually this are RA-accountants with the additional title of RE / CISA (**RA-RE**-accountant).

Regarding the accounting firms in which these types of accountants work as a trusted advisor a distinction can be made into:

AR-firms	: Accounting firms compiling **A**nnual **R**eports
AO-firms	: **A**udit-**O**nly firms, focussing on (IT)audits
MX-firms	: **Mix**ed accounting firms compiling annual reports and performing (IT)audits

MSEs/SMEs usually obtain services of a trusted advisor from an accountant working at either an AR-firm or a MX-firm. Based on the type of accountant (and IT-auditor) in relation to the type of accounting firm the following overview can be made of combinations that are most common:

Table 91: Accountant types at different accounting firms.

	TYPE OF ACCOUNTANT		
TYPE OF ACCOUNTING FIRM	AA	RA	RA-RE
AR-FIRM	☑	☑	□
AO-FIRM	□	☑	☑
MX-FIRM	☑	☑	☑

[11] *Document: SRA_UU_Brainstorm_10_05_2021.docx*

[12] *Accountant consultant (source: nba.nl)*

[13] *Chartered accountant (source: nba.nl)*

[14] *IT auditor (source: norea.nl)*

Based on the first definition of the personas, the type of accountants and accounting firms a clearer image of the personas was defined:

AC-AR	: The accountant (AA or RA) working in a AR-firm
RA-MX	: The RA-accountant working in a MX-firm
RA-RE-MX	: The RA-accountant with IT-knowledge working in a MX-firm

Next to these three personas there is the Trainer / teacher for the Certified Security Defenders training. Note that the certification criteria for the level of Certified Security Defender have not been set yet.

In relation to the current education program[15] in which five Security Defender Levels and Pillars of Expertise are described the three personas can be presented as follows (Figure 49).

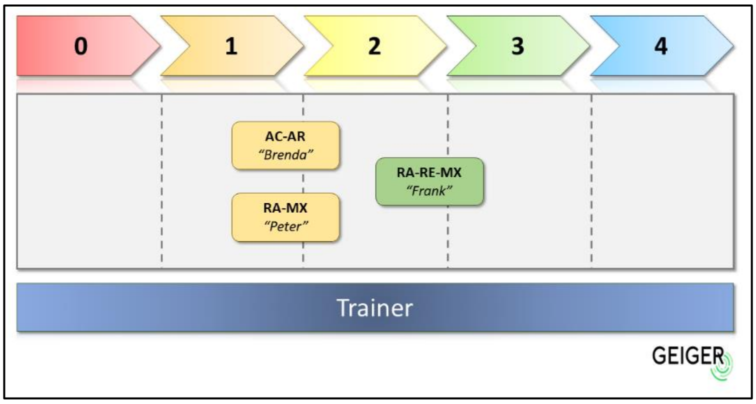


Figure 48: Accountant types and their knowledge levels.

Based on discussion in previous sessions we expect that two personas AC-AR ("Brenda") and RA-MX ("Peter") are on level 1 / 2 and that RA-RE-MX ("Frank") is already partially on level 3. We expect the Trainer to be on level 3 or 4 with training skills. For each of the personas a brief overview of the user journey is defined in Table 92.

Table 92: Accountant journey steps and their descriptions.

User Journey Step	Description
Awareness	Expected level of knowledge of cyber security and expected attitude on the use of the GEIGER application.
Onboarding	Arguments for participation in the project which can be used in communication.
Configuration	Knowledge used to identify gaps and risks at the SME, which help to steer advice on the how to use GEIGER closing this gap.
Scan	Use of the tool, scanning for risks and threats and support needed.
Security help	Who is expected to be called upon for help interpreting the results of the scan?
Feedback	The way of providing advice and feedback towards the SME.

5.2.2.2 Validation Phase

Each of the personas has their own professional seasonal pattern which must be considered. The pattern for AC-AR ("Brenda") and RA-MX ("Peter") differs from that of RA-RE-MX ("Frank"). For the SMEs involved, customers of the accounting firms, it's expected that the impact of this is negligible.

[15] cloud.cyber-geiger.eu/.../WP3 Security Defenders Education/D3.1 Training Plan/GEIGER Security Defende levels pillars.docx

Table 93: Accountant journey steps and the period where they are expected to be performed.

Activity	Persona	Period	Description
Awareness	All	M12 - M30	Creating awareness. Informing accountants / accounting firms about the possibilities of GEIGER.
Onboarding	All	M18 - M30	Onboarding accountants in participating validation of the GEIGER tool. Starting with a kick-off meeting on January 20, 2022.
Education / Training	AR-AC RA-MX RA-RE-MX	M23 - M30 M23 - M30 M20 - M30	Education / training accountants in both the use of the GEIGER tool as well as cyber security in general.
Validation / Pilot	AR-AC RA-MX RA-RE-MX	M23 - M30 M23 - M30 M20 - M30	Use of the tool, scanning for risks and threats and support needed.
Evaluation	All	M20 - M30	Monthly evaluation on the results of both the training and the validation

Note that to be able to provide Education / Training a useable version of the GEIGER-tool must be available. Pending the criteria that must be met to qualify as a CSD.

5.2.3 T4.3: Romania

During the brainstorming session we have identified several personas to the level of digitally dependent micro-enterprises (ME). They are:

- Educator (trainer) [**TR**]
- Certified defender [**CSD**]
- Employee with non-IT background [**EN**]
- Facilitator [**FA**]
- Manager of non-IT start-up [**MA**]
- Technical staff with IT background [**IT**]

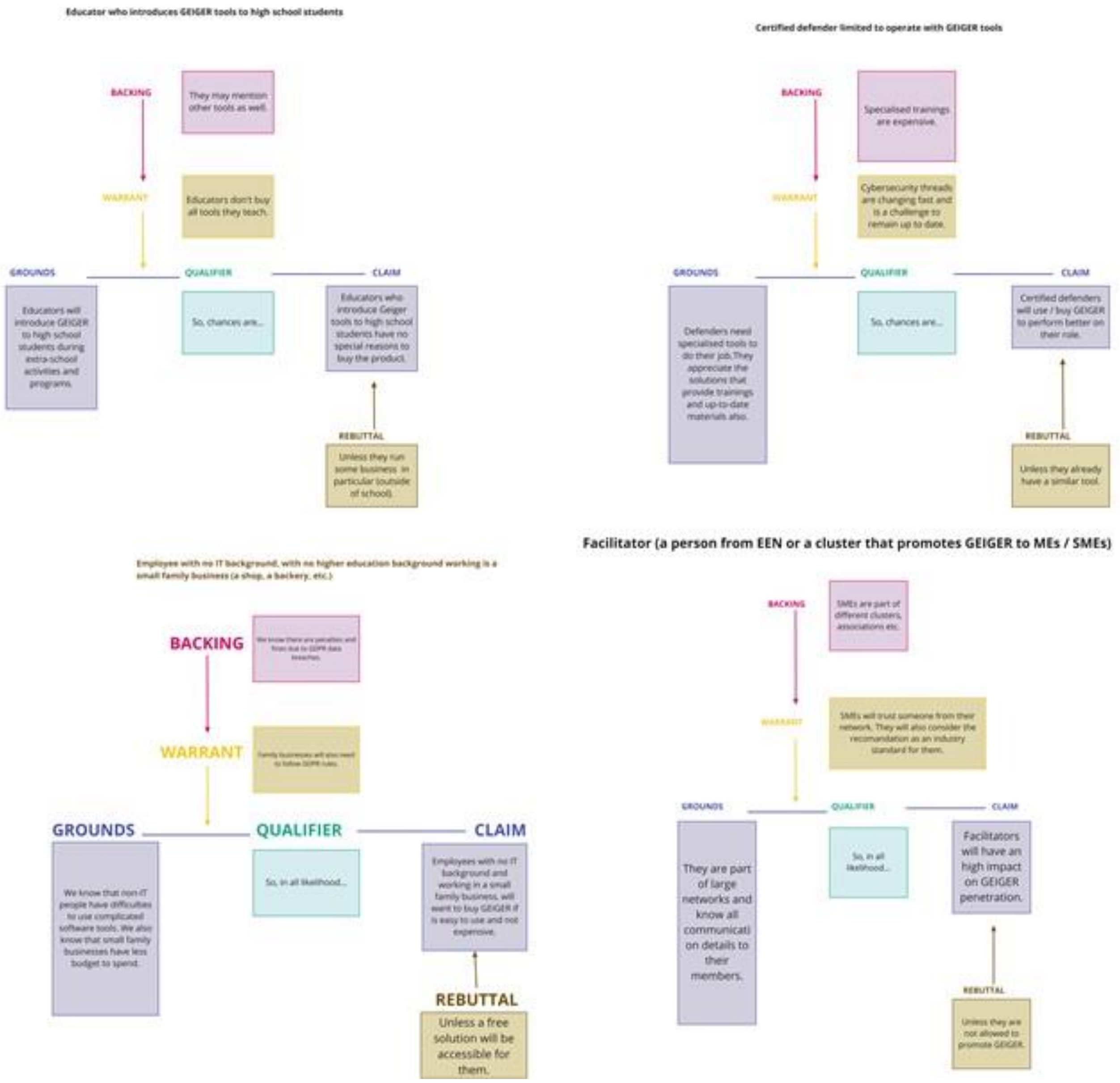
Table 94 illustrates the perspective of persona in the validation phase of the cyber-GEIGER solution.

Table 94: ME persona journey steps and the period where they are expected to be performed.

Activity	Persona	Period	Description
Awareness	All	M12 - M30	Creating awareness. Informing firms about the possibilities of GEIGER.
Onboarding	All	M18 - M30	Onboarding in participating validation of the GEIGER tool. Starting with a kick-off meeting on December 10, 2022.
Education / Training	All	M23 - M30 M23 - M30 M20 - M30	Education / training in both the use of the GEIGER tool as well as cyber security in general.
Validation / Pilot	EN FA MA IT	M23 - M30 M23 - M30 M20 - M30	Use of the tool, scanning for risks and threats and support needed.
Evaluation	EN FA MA IT	M20 - M30	Monthly evaluation on the results of both the training and the validation

Note that to be able to provide an Education / Training useable version of the GEIGER-tool must be available. Pending the criteria that must be met to qualify as a CSD.

Argumentation of each persona is illustrated in Figure 49.


Educator who introduces GEIGER tools to high school students
BACKING
They may mention other tools as well.
WARRANT
Educators don't buy all tools they teach.
GROUNDS
Educators will introduce GEIGER to high school students during extra-school activities and programs.
QUALIFIER
So, chances are...
CLAIM
Educators who introduce Geiger tools to high school students have no special reasons to buy the product.
REBUTTAL
Unless they run some business in particular (outside of school).
Certified defender limited to operate with GEIGER tools
BACKING
Specialised trainings are expensive.
WARRANT
Cybersecurity threads are changing fast and is a challenge to remain up to date.
GROUNDS
Defenders need specialised tools to do their job.They appreciate the solutions that provide trainings and up-to-date materials also.
QUALIFIER
So, chances are...
CLAIM
Certified defenders will use / buy GEIGER to perform better on their role.
REBUTTAL
Unless they already have a similar tool.
Employee with no IT background, with no higher education background working is a small family business (a shop, a backery, etc.)
BACKING
We know there are penalties and fines due to GDPR data breaches.
WARRANT
Family businesses will also need to follow GDPR rules.
GROUNDS
We know that non-IT people have difficulties to use complicated software tools. We also know that small family businesses have less budget to spend.
QUALIFIER
So, in all likelihood...
CLAIM
Employees with no IT background and working in a small family business, will want to buy GEIGER if is easy to use and not expensive.
REBUTTAL
Unless a free solution will be accessible for them.
Facilitator (a person from EEN or a cluster that promotes GEIGER to MEs / SMEs)
BACKING
SMEs are part of different clusters, associations etc.
WARRANT
SMEs will trust someone from their network. They will also consider the recomandation as an industry standard for them.
GROUNDS
They are part of large networks and know all communicati on details to their members.
QUALIFIER
So, in all likelihood...
CLAIM
Facilitators will have an high impact on GEIGER penetration.
REBUTTAL
Unless they are not allowed to promote GEIGER.

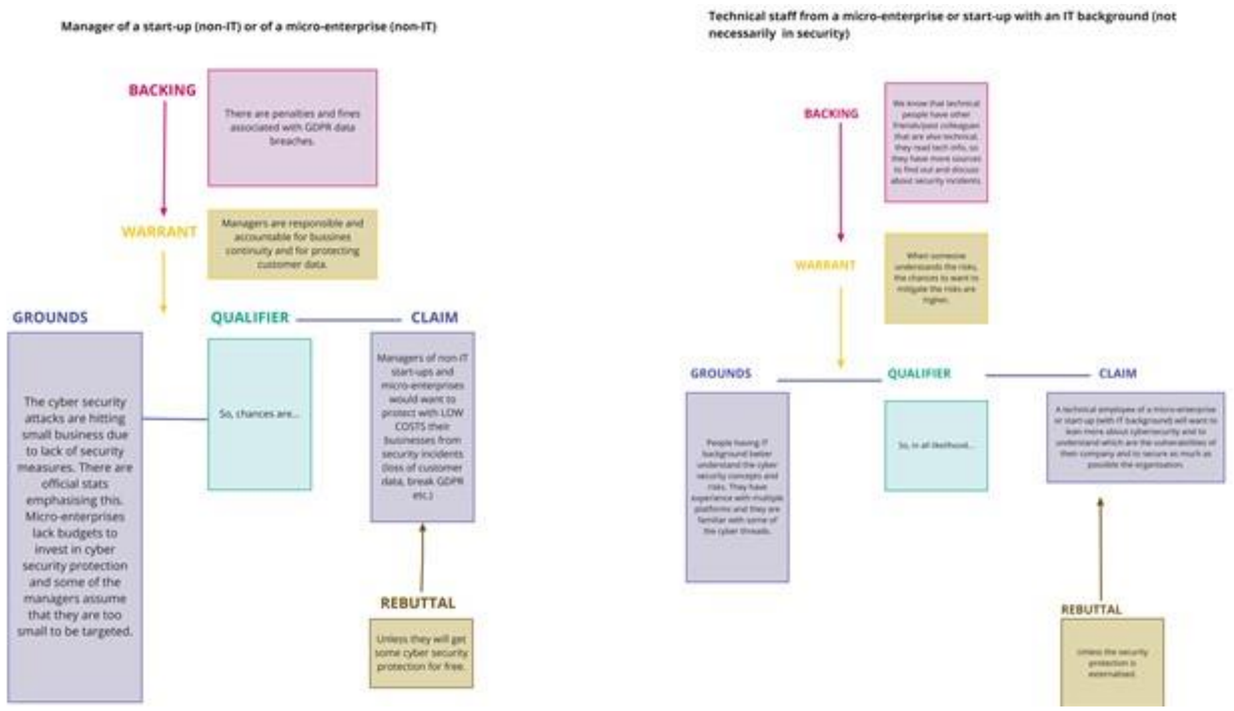

Manager of a start-up (non-IT) or of a micro-enterprise (non-IT)
BACKING
There are penalties and fines associated with GDPR data breaches.
WARRANT
Managers are responsible and accountable for bussines continuity and for protecting customer data.
GROUNDS
The cyber security attacks are hitting small business due to lack of security measures. There are official stats emphasising this. Micro-enterprises lack budgets to invest in cyber security protection and some of the managers assume that they are too small to be targeted.
QUALIFIER
So, chances are...
CLAIM
Managers of non-IT start-ups and micro-enterprises would want to protect with LOW COSTS their businesses from security incidents (loss of customer data, break GDPR etc.)
REBUTTAL
Unless they will get some cyber security protection for free.
Technical staff from a micro-enterprise or start-up with an IT background (not necessarily in security)
BACKING
We know that technical people have other friends/past colleagues that are also technical, they read tech info, so they have more sources to find out and discuss about security incidents.
WARRANT
When someone understands the risks, the chances to want to mitigate the risks are higher.
GROUNDS
People having IT background better understand the cyber security concepts and risks. They have experience with multiple platforms and they are familiar with some of the cyber threads.
QUALIFIER
So, in all likelihood...
CLAIM
A technical employee of a micro-enterprise or start-up (with IT background) will want to lean more about cybersecurity and to understand which are the vulnerabilities of their company and to secure as much as possible the organisation.
REBUTTAL
Unless the security protection is externalised.

Figure 49: Argumentation of persona in the case of ME.

To test and validate the GEIGER solution in the ME environment from Romania, we considered the following action plan:

- Test the first release (basic prototype) with 5 MEs, two being partners in the project, and other three from the Cluj IT ecosystem. Two of MEs are start-ups, the other three have more than five years on the market.
- Use multipliers to enter contact with a wider range of MEs. In this respect, we have contacted 10 Chambers of Commerce and Industry, an Owner Association for Small Businesses and Handcrafts, and a Club of Entrepreneurs of SMEs managed by a bank dedicated for this target group.
- We have started to test the theoretical prototype of the solution (value proposition) and we are now in the process of analyzing results.
- Using the network of multipliers, we informed the contacted MEs and small businesses about the next steps, meaning the training program and the iterative process of testing the next versions of the prototype (MVP, alpha, beta versions).

Our team works close with the coordinator of WP3 to provide feedback and suggestions for the educational program. We scheduled to start the training activities at the beginning of 2022. The basic prototype will be tested in November-December, once we have the solution released. Testing will be focused on user experience, and functionalities. In this respect we will propose a testing plan, agreed with the coordinator of WP4.

6 Data Protection Impact Assessment (DPIA)

In any validation activity involving data collection and processing, we should evaluate whether a data protection impact assessment (DPIA) is necessary. For all activities where the conclusion is that a DPIA is necessary, we perform the DPIA before the start of the validation phase in which that activity is conducted. In Section 6.1 we will first describe the principles driving DPIAs conducted within the GEIGER project. Then, in Section 6.2, we will cover the DPIAs conducted for the first phase of validation. For the later phases of validation, we intend to conduct more extensive DPIAs, the plan for which we present in Section 6.3.

6.1 Principles

In GDPR, DPIAs are mandatory according to Art. 35 if the processing of personal data results in a high risk to the rights and freedoms of natural persons. DPIA is defined as a procedure that describes the processing of personal data by assessing the necessity, proportionality, and risks of processing and defining measures to address these risks. Compared to common risk management procedures, the risk assessment in a DPIA does not focus on the risks of organisations/projects in context with their activities but on the risks of individuals. The assessment is required before a new technology or a new way of data processing is applied. It is important that organisations/projects document that they have considered DPIA and can show evidence that a DPIA was conducted.

Respecting data protection is a high priority of the GEIGER project. In the ethics work package (WP7) and the deliverable D7.2 POPD, we have provided an overview of the data processing and we have elaborated on the data minimisation principle. Additionally, in D7.2 POPD, the measures for obtaining user consent and the responsibilities of the controllers (FHNW, BBB, SRA, and CLUJ-IT) have been defined.

Building upon WP7 guidelines and upon the GDPR, DPIAs for the data processing as part of the GEIGER validation in WP4 has been conducted. This includes in particular

- The assessment of data processing activities in initial phase (within the consortium)
- The assessment of data processing activities in subsequent validation phases (beyond the project consortium). Hence, the assessment is referring to each individual GEIGER use case (Switzerland, Romania, and The Netherlands).

The DPIA assessment is documented and supported by the Excel tool that has been developed in scope of the GEIGER project. FHNW is serving as a moderator, while the WP4 leader and the WP4 task leaders are the ones responsible for defining and assessing the risks of the data processing as well as for determining mitigating actions.

6.2 Implementation for Validation Phase 1

Table 95 provides a brief overview of the activities conducted during the first phase of validation. For the definition of the GEIGER support function, no data is collected. For various other activities, some data may be collected, but it is not of a personal nature and often largely involves data provided by partners within the GEIGER consortium.

Three activities do involve a data protection risk. The first involves action research performed with the GEIGER use case partners. The collected data will involve observations on how users use GEIGER and will necessarily involve the recording of personal data to link observations to specific GEIGER persona for validation. The DPIA for this activity also encompasses the activities performed for testing the GEIGER UI.

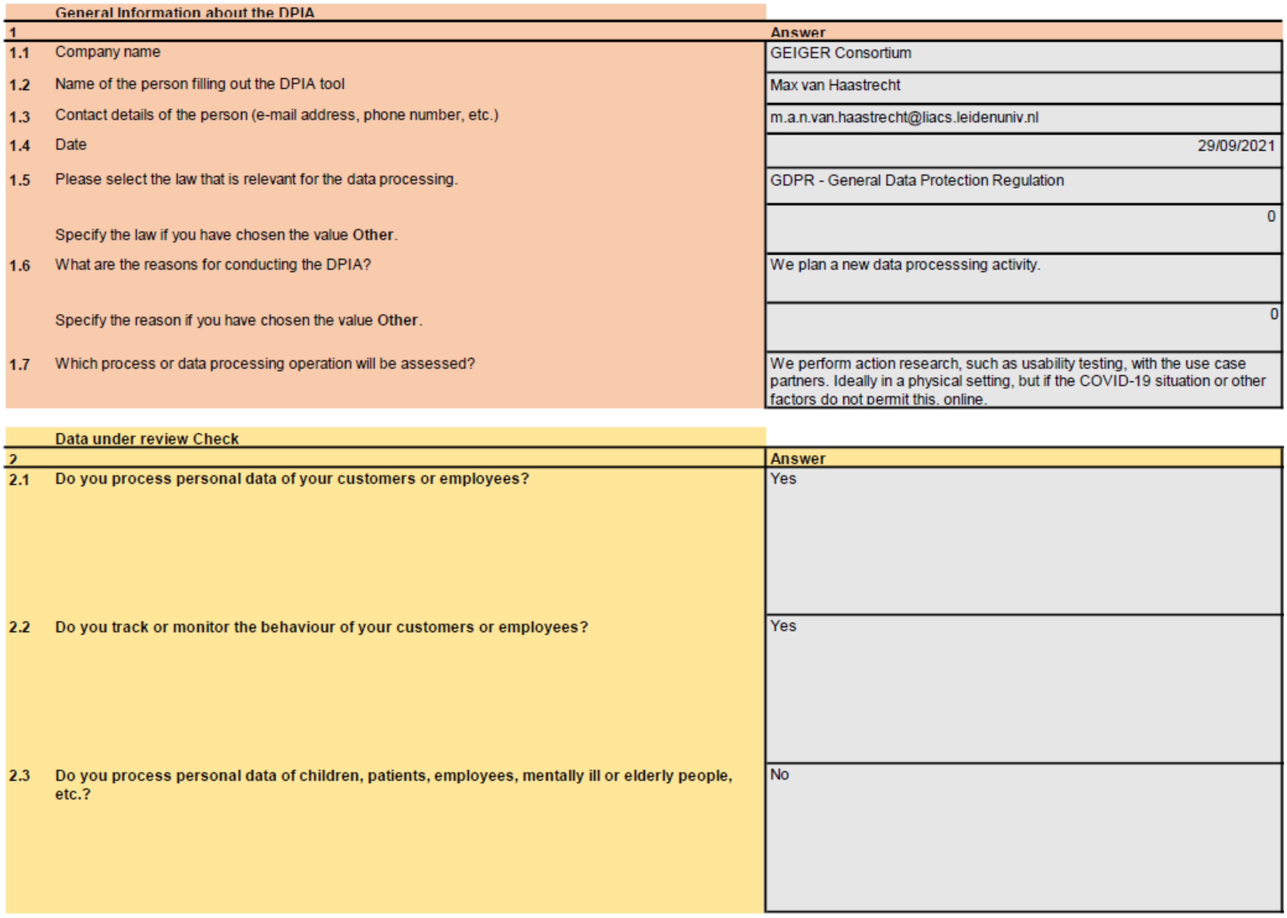
The GEIGER multiplier survey involves the collection of personal data, albeit from partners in the GEIGER consortium. Additionally, the collected data is not of a sensitive nature, which is why this data processing activity is classed as ‘medium risk.’ Finally, the GEIGER use case partner survey is classed as ‘high risk.’ The survey not only involves the collection of personal data, but this data may also be sensitive as it pertains to the abilities of users and their MSE (awareness, resilience, etc.), as well as to the occurrence of incidents at their MSE.

Table 95: Activities of validation phase 1, along with an indication of their data processing risk.

Activity	Data Collection
Expert evaluation of GEIGER content	No risk associated with collected data.
GEIGER support definition and creation	No data collected.
Recruit GEIGER alpha users	No risk associated with collected data.
Technical experiments to evaluate solution	No risk associated with collected data.
Action research with use case partners	High risk, meaning a DPIA is mandatory, and should be extensively motivated.
GEIGER multiplier survey (monthly)	Medium risk, meaning a DPIA is desirable.
GEIGER use case partner survey (monthly)	High risk, meaning a DPIA is mandatory, and should be extensively motivated.

For the activities identified as having some level of risk, we performed a DPIA using the 'Data Protection Impact Assessment Tool' of FHNW, as described in deliverables D1.1 'Requirements' and D6.2 'Y1 and Periodic Report'. The Excel tool provides a structured template to address GDPR concerns related to data processing activities, eventually guiding the user to produce their own set of measures to reduce risks.

The DPIA tool begins by asking some basic questions on the person filling in the DPIA and the data processing activity being assessed. Figure 50 shows the provided answers for the case of action research with use case partners.

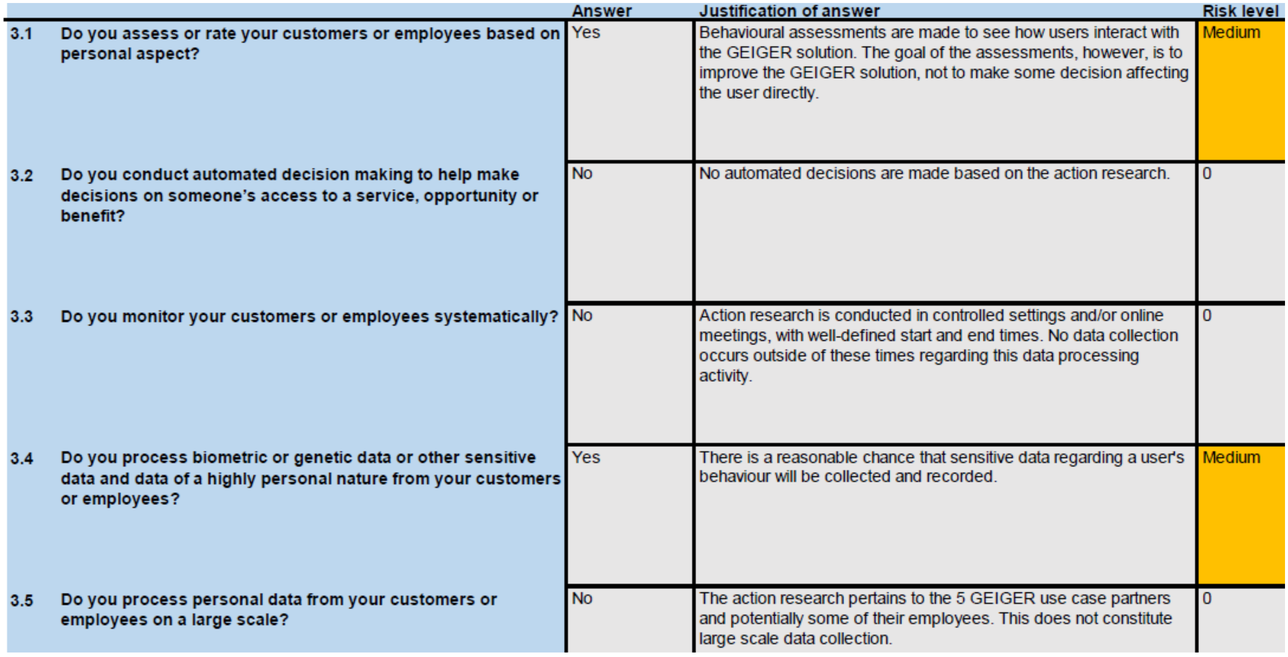


	General Information about the DPIA	
1		Answer
1.1	Company name	GEIGER Consortium
1.2	Name of the person filling out the DPIA tool	Max van Haastrecht
1.3	Contact details of the person (e-mail address, phone number, etc.)	m.a.n.van.haastrecht@liacs.leidenuniv.nl
1.4	Date	29/09/2021
1.5	Please select the law that is relevant for the data processing.	GDPR - General Data Protection Regulation
	Specify the law if you have chosen the value Other.	0
1.6	What are the reasons for conducting the DPIA?	We plan a new data processsing activity.
	Specify the reason if you have chosen the value Other.	0
1.7	Which process or data processing operation will be assessed?	We perform action research, such as usability testing, with the use case partners. Ideally in a physical setting, but if the COVID-19 situation or other factors do not permit this, online.

	Data under review Check	
2		Answer
2.1	Do you process personal data of your customers or employees?	Yes
2.2	Do you track or monitor the behaviour of your customers or employees?	Yes
2.3	Do you process personal data of children, patients, employees, mentally ill or elderly people, etc.?	No

Figure 50: Initial questions and answers for the DPIA tool.

Once these initial questions have been filled in, one proceeds to answer more detailed data processing questions to uncover any potential data processing risks related to GDPR. Figure 51 shows a selection of the answers related to our action research with use case partners. We observe that several questions yield answers implying a medium risk level. These risks should be either mitigated or reduced by corresponding measures.



		Answer	Justification of answer	Risk level
3.1	**Do you assess or rate your customers or employees based on personal aspect?**	Yes	Behavioural assessments are made to see how users interact with the GEIGER solution. The goal of the assessments, however, is to improve the GEIGER solution, not to make some decision affecting the user directly.	Medium
3.2	**Do you conduct automated decision making to help make decisions on someone's access to a service, opportunity or benefit?**	No	No automated decisions are made based on the action research.	0
3.3	**Do you monitor your customers or employees systematically?**	No	Action research is conducted in controlled settings and/or online meetings, with well-defined start and end times. No data collection occurs outside of these times regarding this data processing activity.	0
3.4	**Do you process biometric or genetic data or other sensitive data and data of a highly personal nature from your customers or employees?**	Yes	There is a reasonable chance that sensitive data regarding a user's behaviour will be collected and recorded.	Medium
3.5	**Do you process personal data from your customers or employees on a large scale?**	No	The action research pertains to the 5 GEIGER use case partners and potentially some of their employees. This does not constitute large scale data collection.	0

Figure 51: DPIA tool data processing questions.

For each of the three validation activities for which we conducted a DPIA, we will list the elements where a medium or high risk level was constituted and name our mitigation measures. Table 96 lists the data processing dimensions for our action research activities with use case partners. In the case of this activity, each element was first classed as having medium risk, and after the mitigation measure(s) reducing the risk was classed as having low risk.

Table 96: Medium-risk data processing dimensions for our action research with use case partners.

Question	Explanation	Measure(s)
Do you assess or rate your customers or employees based on personal aspects?	Behavioural assessments are made to see how users interact with the GEIGER solution. The goal of these assessments, however, is to improve the GEIGER solution, not to make some decision affecting the user directly.	Consent forms will be provided before any personally identifiable behaviour is recorded. Additionally, use case partners will be provided with all information provided by and related to them during the action research sessions. They will have the ability to choose to delete any information they find too sensitive.
Do you process biometric or genetic data or other sensitive data and data of a highly personal nature from your customers or employees?	There is a reasonable chance that sensitive data regarding a user's behaviour will be collected and recorded.	Anyone involved in recording data related to this data processing activity will be instructed to avoid noting any highly personal elements and unnecessarily sensitive data. Additionally, users will receive a consent form beforehand where they give permission to record data. Lastly, users will be presented with all collected and recorded data relating to them and be given the chance to delete any data they wish to.

Do you match, compare or combine data records from your customers or your employees from multiple sources?	The research conducted here will likely span several session guided by different people. Where one person may be responsible for usability testing in online click dummy sessions, another may be responsible for user testing in a physical setting. This data will need to be combined to form a complete picture.	All persons involved in data collection will be given extensive instructions on which data processing steps are relevant to the goal of this activity. All data involved in this data processing activity must be stored on the GEIGER cloud and only on the GEIGER cloud, unless being temporarily used for local analysis by those persons identified as being allowed to perform this analysis. These persons will be listed in Deliverable D4.1. Consent forms and communication to users will explicitly mention which activities belong to this data processing activity, so they are aware which data may be combined.
Do you rate or score your customers or employees?	Data will be used to analyse, correlate, and assess behavioural patterns.	Any data files collected for this activity must be explicitly and clearly marked using the Traffic Light Protocol (TLP) classification as defined for GEIGER. No data that is made publicly available (for example as part of deliverables or other publications) should contain any personally identifiable information, or information allowing to identify a specific company or set of companies.

For the GEIGER multiplier survey, we conducted a DPIA since it was classed among the activities requiring a DPIA. However, we found only one potential risk during this process, and this was considered a low risk. This meant no further actions had to be taken, although a consent form was nevertheless used for the survey and we still treat the data carefully by allowing only access to those who strictly need it.

The risk in question related to the element: Do you rate or score your customers or employees? In the multiplier survey we do not rate or score in a traditional sense, but we do use the data to assess whether particular multipliers should be pushed more to attract more alpha and/or beta users. For us, this constitutes an accepted risk in terms of data protection. We have also actively communicated to the multipliers that this is our intended purpose of conducting the survey, so they are aware of what is intended with the provided data before filling in the survey.

Table 97 shows the medium-risk data processing dimensions for our use case partner survey. Besides these two medium risks, which are reduced to low risks by our mitigation measures, we again encountered a low risk regarding the 'do you rate or score your customers or employees?' question. The data collected for the survey will be analysed to find correlations regarding GEIGER use and improvements in the cybersecurity domain. However, this is purely intended for use in improving the GEIGER solution itself. Hence, given the consent form we provide to the use case partners before filling in the surveys, we consider this an accepted risk.

Table 97: Medium-risk data processing dimensions for our use case partner survey.

Question	Explanation	Measure
Do you assess or rate your customers or employees based on personal aspect?	We collect data from the use case partners on their opinion of the GEIGER solution. We do this to improve the GEIGER solution, which is the main goal with the survey results. However, we also ask users to provide some	We will limit access to survey data only to those people who require it for analysis and who have received extensive instructions on how they are allowed to process the data. Additionally, use case partners will always first give consent for

	information on their own knowledge level and their company's cybersecurity status, to assess whether GEIGER is achieving its goals of improving awareness and resilience.	processing the data before participating in the survey.
Do you process biometric or genetic data or other sensitive data and data of a highly personal nature from your customers or employees?	We do not collect any highly personal data in the survey, with the most personal data being the name of the person. However, we do collect sensitive data, as we ask people to describe their cybersecurity posture and whether their company experienced a cybersecurity incident in the past month. It is possible that in open questions people also reveal sensitive data beyond what we are asking for.	The collected data will be stored within a centrally managed survey tool at FHNW. A limited number of persons will have access to the data and they will be instructed on how it can be used. Any data processed further, for example for publishing purposes, must be anonymised first. Additionally, we will guide participants to avoid inserting names and other indications to individuals when describing sensitive information related to incidents (data minimisation). We will be transparent to participants as to the usage of questionnaire data.

Since many of the interactions during the first phase of validation are with partners within the consortium, we do not encounter many medium-level risks, and do not encounter any high-level risks. This is likely to change for the later validation phases, where we process data of partners outside of the GEIGER consortium. At times, this may even include data from people under the age of 18 (although always at least of age 16), or data from people that may in some sense be considered vulnerable. For these situations, a more extensive DPIA process may be necessary, to ensure that we deal with privacy and data protection issues adequately. Our intended approach regarding DPIAs for the later validation phases is described in the next section.

6.3 Plan for Validation Phase 2 and Phase 3

In validation Phases 2 and 3, we will increasingly interact with GEIGER users external to the consortium. Additionally, with the initiation of security defender education, we will also begin to collect data from people under the age of 18 (although always 16+). This implies that our DPIA process as implemented for validation Phase 1 is no longer sufficient, and we need to formulate a more rigorous approach to ensure we adequately address any potential data protection concerns.

For validation Phases 2 and 3 we will complement the DPIA process of validation Phase 1 with additional input from the leaders in the GEIGER use case countries: BBB (Switzerland), SRA (The Netherlands), and CLUJ-IT (Romania). The use case country leaders have a better understanding of the needs of those participating in local validation trials, meaning their view is important to consider while conducting Phase 2 and 3 DPIAs.

The DPIA process for Phases 2 and 3 will work as follows:

- We will first define all activities in an upcoming phase where a DPIA is necessary, as we did for Phase 1.
- In the month before the start of the next validation phase, we will meet separately with each use case country leader to walk through the DPIA process. They can provide their view on how they envision local implementations of validation activities, and which consequences this may have in terms of data protection requirements.
- Following the individual country meetings, we will record the results and formalize the selected mitigation measures. We will present the results and measures to all use case country leads, who can provide final feedback on the plan.

- Finally, we will incorporate the use case country lead feedback in our DPIAs and conclude our DPIA process.

We believe that this extended process provides the necessary basis for dealing with the additional DPIA requirements faced in validation Phases 2 and 3.

7 Current State of Validation and KPIs

Although we are only at the outset of the GEIGER validation efforts, it is worth presenting a summary of where we currently are and discussing where we want to be in a years' time as the GEIGER project draws to a close. In this section, we will firstly discuss the KPIs within the GEIGER project that fall under the responsibility of WP4 (Section 7.1). Then, we will link the solution requirements unearthed in WP1 and WP3 to our validation activities in Section 7.2. This provides us with a concise overview of when we can expect feedback related to specific GEIGER features. Finally, we will delineate our view on the result of our eventual validity assessment in Section 7.3

7.1 KPI Measurement

Table 98 lists the KPIs of the GEIGER project that fall under the responsibility of WP4. There is currently not much progress to report regarding the KPIs, as the first months of validation were mostly focused on preparatory work. Nonetheless, we have made considerable progress in ensuring the necessary measures are in place for us to meet these KPIs.

As we outlined in Section 3, we have coupled the GEIGER KPIs directly to our validation approach. This does not only apply to the WP4 KPIs outlined here, but in fact to all KPIs of the GEIGER project. This ensures that all activities in past and future are geared towards gathering the necessary data to measure our KPIs. Just measuring KPIs does not ensure meeting KPIs, but by communicating our results at an early stage, we enable other work packages to take the necessary actions to guarantee a smooth progression.

Additionally, there are several KPIs where we can report significant results. KPI 5.1 states that the 'GEIGER Framework will have been evaluated in ≥ 50 MSEs'. Our local efforts in alpha user recruitment have already resulted in the commitment of over 25 alpha user MSEs. Regarding the educational KPIs related to WP4 (e.g., I2.1.1.3 and I2.1.1.5), local kick-off events in Switzerland and The Netherlands in month M18 produced promising results. Lastly, with the first use case partner surveys initiated and alpha user surveys soon to follow, we have started to make our first inroads for survey related KPIs, such as I2.1.1.9 and KPI I2.1.2.1.

Table 98: KPIs under the responsibility of WP4.

KPI	Description
1.2	Understanding the GEIGER Indicator by CEOs of MSEs ≥ 4.0 on the Mean Opinion Score scale ranging from 1 (bad) to 5 (excellent)
1.3	Predicted attack intensity (e.g., affected devices) for a specific attack matches +/- 20% in more than 60% of all observed time spans retrospectively
2.7	≥ 4.0 Mean Opinion Score concerning trust towards the toolbox.
5.1	3 "Certified Security Defenders" education approaches validated and demonstrated
5.2	GEIGER Framework will have been evaluated in ≥50 MSEs
5.3	Satisfaction by MSEs using the GEIGER Framework ≥ 4.0 on "Mean Opinion Score" scale ranging from 1 (bad) to 5 (excellent)
I2.1.1.3	>50 MSEs will have benefitted from the Security Defender education in the Swiss pilot performed with apprentices by the school BBB
I2.1.1.4	>50 start-up MSEs will have benefitted from the Security Defender education in the Romanian pilot performed by the incubator/accelerator CLUJ IT CLUSTER
I2.1.1.5	>370 MSEs will have benefitted from advice by an accountant with Security Defender education in the Dutch pilot performed by the education provider UU
I2.1.1.6	>50 schools with vocational training for apprentices will intend to adopt the Security Defender education programme at the end of the project
I2.1.1.7	>50 incubators/accelerators will intend to adopt the Security Defender education programme at the end of the project

I2.1.1.8	>50 accountants education providers will intend to adopt the Security Defender education programme at the end of the project
I2.1.1.9	GEIGER capacity-building assessed in surveys with >1’000 responses
I2.1.1.10	GEIGER capacity-building refined in >10 events targeting MSEs.
I2.1.2.1	perceived level transparency of risks ≥ 4.0 on MOS scale
I2.1.2.2	perceived level of decision support for risk reduction ≥ 4.0 on MOS scale
I2.1.2.3	perceived level of risk explanation ≥ 4.0 on MOS scale
I2.1.3.1	at least 80% of basic recommendations for human error prevention are adopted by the pilot MSEs
I2.1.3.2	Shield tools are available and in use by the pilot MSEs for protecting against at least 80% of attacks recommended for protection by the participating CERTs/CSIRTs.
I2.1.3.3	90% of the incidents experienced by the pilot MSEs are detected and resolved within 30 days

7.2 Meeting Requirements

Deliverables D1.1 on Requirements and D3.1 on the GEIGER Training Plan presented various feature requirements for the GEIGER solution. We distinguish three types of features: cloud features, toolbox features, and educational features. Where we have an explicit coupling between KPIs and our validation approach, this is not the case for feature requirements. It is therefore justified to question whether our validation activities adequately address the need to measure whether we are meeting our feature requirements. We will address this question in this section.

Table 99 lists all feature requirements along with the validation activities they can be associated with. The index of the feature requirement indicates whether it corresponds to a cloud feature (C), a toolbox feature (T), or an educational feature (E). All cloud and toolbox features and their indices can be found in Deliverable D1.1. The educational features are included in Deliverable D3.1, although they are less explicitly listed. The only feature requirements we exclude are the educational feature requirements related to specific use case countries, since these requirements are yet to be clarified as we move towards more detailed experimentation in the coming period.

What we can conclude from the table is that each feature requirement is connected to at least one validation activity. We can therefore conclude that we have the necessary validation activities in place to evaluate all feature requirements. Nevertheless, simply noting these connections does not resolve us of the responsibility to explicitly evaluate whether the GEIGER solution’s features adequately cover the requirements. Future validation activities must keep the requirements of Table 99 in mind when designing experiments.

Table 99: GEIGER feature requirements mapped to validation activities.

Index	Feature Requirement	Connected Activities
C.F01	GEIGER Indicator and Recommendations	Action research with use case partners (Phase 1)
C.F01.1	Competent CERT Selection	GEIGER beta user survey (monthly) (Phase 3), Interviews with CERTs to evaluate incident reporting (Phase 3)
C.F01.2	Relevant Industry Selection	Action research with use case partners (Phase 1)
C.F02	Community Profiling	Technical experiments analyzed using statistical methods (Phase 2)
C.F02.1	Cloud Account	Repeated technical experiments to evaluate solution (Phase 2)
C.F02.2	MSE Profile Sync	Repeated technical experiments to evaluate solution (Phase 2)
C.F02.3	Community Analysis	Technical experiments analyzed using statistical methods (Phase 2)
C.F03	Risk Knowledge Base	Expert evaluation of GEIGER content (Phase 1), Interviews with CERTs to evaluate incident reporting (Phase 3)

C.F03.1	Risk Knowledge Curation	Repeated technical experiments to evaluate solution (Phase 2), Formulate a GEIGER business plan (Phase 3)
C.F04	Incident Reporting	Final technical experiments (Phase 3), Interviews with CERTs to evaluate incident reporting (Phase 3)
C.F05	Certified Security Defenders Directory	Final technical experiments (Phase 3), GEIGER CSD Survey (monthly) (Phase 3)
C.F21	CERT Account Management	Interviews with CERTs to evaluate incident reporting (Phase 3)
C.F22	Threat Communication	Technical experiments analyzed using statistical methods (Phase 2), Interviews with CERTs to evaluate incident reporting (Phase 3)
C.F23	Incident Notification	Technical experiments analyzed using statistical methods (Phase 2), Interviews with CERTs to evaluate incident reporting (Phase 3)
C.F41	Management of the Risk Knowledge Base	Action research with use case partners (Phase 1), GEIGER use case partner survey (monthly) (Phase 1), Technical experiments analyzed using statistical methods (Phase 2)
C.F42	Management of the Community Knowledge Base	Repeated technical experiments to evaluate solution (Phase 2), Technical experiments analyzed using statistical methods (Phase 2)
C.F43	Content Curation	Action research with pilot users (Phase 2), GEIGER pilot user survey (monthly) (Phase 2)
C.F44	Management of the Certified Security Defenders Directory	Final technical experiments (Phase 3), GEIGER CSD Survey (monthly) (Phase 3), Internal expert evaluation of CSD registration and updating (Phase 3)
T.F01	Toolbox Installation	Action research with use case partners (Phase 1), Action research with pilot users (Phase 2)
T.F01.1	Toolbox Updating	Technical experiments analyzed using statistical methods (Phase 2), Action research and interviews with tool owners (Phase 3)
T.F01.2	Device Pairing	Technical experiments analyzed using statistical methods (Phase 2), Final technical experiments (Phase 3)
T.F01.3	Cloud Account Pairing	Technical experiments analyzed using statistical methods (Phase 2), Final technical experiments (Phase 3)
T.F01.4	Employee Account Pairing	Action research with pilot users (Phase 2), Technical experiments analyzed using statistical methods (Phase 2), Final technical experiments (Phase 3)
T.F02	MSE Profiling	Technical experiments analyzed using statistical methods (Phase 2), GEIGER beta user survey (monthly) (Phase 3)
T.F02.1	Questionnaire	Expert evaluation of GEIGER content (Phase 1), Technical experiments to evaluate solution (Phase 1), Action research with use case partners (Phase 1)
T.F02.2	Scanner	Technical experiments to evaluate solution (Phase 1), Repeated technical experiments to evaluate solution (Phase 2), Action research and interviews with tool owners (Phase 3)
T.F02.3	Education Reporting	Educational case study (Phase 2)
T.F03	GEIGER Indicator and Recommendations	Action research with use case partners (Phase 1), GEIGER beta user survey (monthly) (Phase 3)
T.F04	Asset Protection	Final technical experiments (Phase 3), Action research and interviews with tool owners (Phase 3)
T.F04.1	Cybersecurity Tool Installation	Final technical experiments (Phase 3), Action research and interviews with tool owners (Phase 3)
T.F04.2	Software Configuration	GEIGER pilot user survey (monthly) (Phase 2), Action research with pilot users (Phase 2)
T.F04.3	Employee Education	Expert evaluation of education and training plan (Phase 2), Educational case study (Phase 2)

T.F05	Incident Reporting and Resolution Guidance	Action research with use case partners (Phase 1), GEIGER use case partner survey (monthly) (Phase 1), GEIGER pilot user survey (monthly) (Phase 2), Interviews with CERTs to evaluate incident reporting (Phase 3)
T.F05.1	Incident Notification	Action research and interviews with tool owners (Phase 3), Interviews with CERTs to evaluate incident reporting (Phase 3), GEIGER beta user survey (monthly) (Phase 3)
T.F06	Data Management	GEIGER pilot user survey (monthly) (Phase 2), GEIGER beta user survey (monthly) (Phase 3)
T.F06.1	Dynamic Consent	Expert evaluation of GEIGER content (Phase 1), GEIGER use case partner survey (monthly) (Phase 1)
T.F07	Threat Updates	Repeated technical experiments to evaluate solution (Phase 2), Technical experiments analyzed using statistical methods (Phase 2)
E.F01	Experiential Learning	Subject-based experiments and survey with alpha (C)SDs (Phase 2), Educational case study (Phase 2)
E.F01.1	Game-Based Learning	Subject-based experiments and survey with alpha (C)SDs (Phase 2), Educational case study (Phase 2)
E.F02	Reverse Mentoring	Subject-based experiments and survey with alpha (C)SDs (Phase 2), Educational case study (Phase 2), GEIGER CSD survey (monthly) (Phase 3)
E.F04.1	MSE Specific	Expert evaluation of GEIGER content (Phase 1)
E.F04.2	GEIGER-Related Topics	Expert evaluation of GEIGER content (Phase 1), Expert evaluation of education and training plan (Phase 2)
E.F04.3	Train-the-Trainer-Schemes	Expert evaluation of education and training plan (Phase 2), GEIGER educator survey (monthly) (Phase 2)
E.F05.1	Curricular Interoperability with EDU Providers	Expert evaluation of education and training plan (Phase 2), GEIGER educator survey (monthly) (Phase 3)
E.F05.2	Technical Interoperability with Toolbox	Repeated technical experiments to evaluate solution (Phase 2), Expert evaluation of education and training plan (Phase 2), Technical experiments analyzed using statistical methods (Phase 2)

7.3 Validity Assessment

Although we discussed the topic of validity at length in Section 3, it is worth reiterating how we view the eventual validity assessment which is the output of WP4 work. One remark we wish to make is that our validation efforts will never result in an overall judgement that the GEIGER project is 'valid' or 'invalid.'

In the argument-based approach to validation we form the GEIGER interpretation and use argument (IUA) chain and address any raised rebuttals with the necessary warrants and backings. In links of our argumentation chain that are determined to be 'weak', more warrants and backings will be required. In links of our argumentation chain that have a high level of face validity, less extensive argumentation is required to 'strengthen' our link.

In Section 5.1.1, we presented and discussed the results of the first of a series of expert panels that will be conducted during validation. We identified areas where our validation efforts should be focused, and areas where the GEIGER solution seems to be making a strong case for validity already. The UI testing results of Section 5.1.2 similarly showed that the current GEIGER solution has its strengths and weaknesses. It is up to WP4 to clearly communicate any potential weaknesses to other work packages, to ensure timely corrective action.

In Section 2, we presented a series of arguments why basing our validity approach purely on achieving KPIs and meeting requirements is not sufficient. Similarly, our validity assessment will not be purely based on an assessment of KPIs and requirements. That said, it is our task in WP4 to continuously monitor progress regarding KPIs and requirements and steer the project in the right direction where necessary.

In the end, we aim to strengthen our argumentation chain in the coming 12 months. When even the most critical of GEIGER users cannot break our argumentation chain, we have succeeded in reaching our validation objectives, and, more importantly, have succeeded in creating a successful application to help MSEs throughout Europe become more cyber resilient.

7.4 COVID-19 Impact

The GEIGER project started during the COVID-19 pandemic. The pandemic has influenced many activities within the project and will likely continue to do so until the end of the project in the winter of 2022. Given the interactive nature of many of the activities we plan to conduct in WP4, we should not underestimate the impact that extended lockdowns and other preventive measures will have on our progress.

Besides the many activities that have already taken place online (e.g., user interface testing), we have also had our first cancellation of a planned physical event: the Dutch kick-off. Our first phase of validation involved mainly the GEIGER use case partners, who have shown a great willingness to participate in online validation sessions where necessary. However, as we start to move to our alpha and beta validation phases, with users who are less intimately connected to the GEIGER project, it will become increasingly difficult to perform all WP4 activities online.

We have mentioned before that the modular structure of our validation planning allows us to be flexible where necessary. Nevertheless, we are dependent on the GEIGER launch event and other local kick-off events to recruit enough users for validation. If these events continue to be cancelled due to the COVID-19 situation, we will need to be creative as to how we recruit new users, and realistic as to our expectations of the number of users reached.

One measure we will implement is to always have a virtual or hybrid back-up plan for any offline event. Of course, we hope to organize as many physical events as possible in 2022. Yet, the reality is that it will remain uncertain for the foreseeable future whether such events can realistically take place. We felt an obligation to acknowledge the limitations the COVID-19 pandemic places on the validation efforts of WP4. Nonetheless, we are confident that we have the necessary measures in place to successfully validate the GEIGER solution.

8 Summary and Conclusion

GEIGER's validation activities span the project months M13 to M30, with M30 also being the final month of the project timeline. In this deliverable we have described the work carried out over the first six months of validation. We have described how the goals of the GEIGER project drive our holistic validation objective, where: We aim to validate whether GEIGER achieves its goals and meets user requirements in operational environments.

In our validation approach we intend to blend the strive for relevance in the GEIGER project with a strive for rigor. With this idea in mind, we introduced a theoretical validation framework inspired by a validation framework for formative assessment from the educational measurement field. Our validation framework provides a firm theoretical basis for us to perform our practical validation activities. It gives us the necessary structure to enable many validation activities in three use case countries spanning across Europe, while maintaining oversight of the progress made and the validity arguments elicited.

Another major result achieved in the first months of our validation work package, is the translation of our theoretical validation framework into a practical validation planning. We divided our validation approach into three phases, yielding a flexible approach that can adapt to unforeseen changes in the planning of the overall GEIGER project.

In Section 5, we described how we have already turned the plans for the first phase of validation into action. General activities covering all use case countries have been complemented with specific activities addressing the needs of the different personas we are dealing with in the GEIGER project. Together, the combination of a top-down and bottom-up approach to validation allows us to answer a variety of validation questions we may otherwise not have been able to answer.

Section 6 covered our process related to data protection impact assessments (DPIAs). We place the utmost importance on the protection of the data processed during validation and the preservation of the privacy of those whose data we collect in WP4. In future data processing activities, especially those dealing with vulnerable persons, we will continue to place these principles at the top of our agenda.

We observed in Section 7 that although we have made considerable progress on both the theoretical and practical sides of validation, much is still left to be done. Heavy collaboration with the other work packages of the GEIGER project needs to continue to ensure we measure and meet our KPIs and requirements. Nevertheless, we need to remain objective and critical within the validation work of WP4. The other work packages of the GEIGER consortium have made impressive progress in creating a relevant solution for MSEs throughout Europe; it is our task to test the validity of their ideas and complement the relevance of GEIGER with a layer of rigor.